

(2)

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION

AD-A211 260

LE

R(S)

1b RESTRICTIVE MARKINGS

FILE COPY

3 DISTRIBUTION/AVAILABILITY OF REPORT

Approved for public release; distribution unlimited.

5 MONITORING ORGANIZATION REPORT NUMBER(S)

AFOSR-88-0140

6a NAME OF PERFORMING ORGANIZATION

New York University

6b OFFICE SYMBOL
(if applicable)

7a NAME OF MONITORING ORGANIZATION

Air Force Office of Scientific Research/NL

6c ADDRESS (City, State, and ZIP Code)

Psychology Department
New York, NY 10003

7b ADDRESS (City, State, and ZIP Code)

Building 410
Bolling AFB, DC 20332-64488a NAME OF FUNDING/SPONSORING
ORGANIZATION

AFOSR

8b OFFICE SYMBOL
(if applicable)

NL

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

AFOSR-88-0140

8c ADDRESS (City, State, and ZIP Code)

Building 410
Bolling AFB, DC 20332-6448

10 SOURCE OF FUNDING NUMBERS

PROGRAM
ELEMENT NO.
61102FPROJECT
NO.
2313TASK
NO.
A5WORK UNIT
ACCESSION NO.

11 TITLE (Include Security Classification)

Kinetic Depth Effect and Optic Flow I. 3D Shape From Fourier Motion

12 PERSONAL AUTHOR(S)

George Sperling

13a TYPE OF REPORT

Reprint

13b TIME COVERED

FROM TO

14 DATE OF REPORT (Year, Month, Day)

1987

15 PAGE COUNT

70

16 SUPPLEMENTARY NOTATION

17 COSATI CODES

FIELD

GROUP

SUB-GROUP

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

DTIC
ELECTE
AUG 11 1989

D

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT

☒ UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT ☐ DTIC USERS21 ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL

Dr. John F. Tangney

22b TELEPHONE (Include Area Code)

(202) 767-5021

22c OFFICE SYMBOL

NL

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted.
All other editions are obsolete.SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

89

6 1 5

AFOSK-IR. 89-1037

Mathematical Studies in Perception and Cognition

88 - 3

Kinetic Depth Effect and Optic Flow

I. 3D Shape from Fourier Motion

Barbara A. Doshier,^{*} Michael S. Landy,^{**}
and George Sperling^{**}

^{*} Psychology Department
Columbia University

^{**} Psychology Department
New York University

Running Head: KDE and Optic Flow I

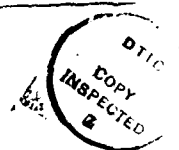
Keywords:

Kinetic Depth Effect
Structure from Motion
Shape Identification
Fourier Motion

Address correspondence to:

Barbara A. Doshier
Department of Psychology
Box 28 Schermerhorn Hall
Columbia University
New York, New York 10027

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Special
A-1	



ABSTRACT

Fifty-three different 3D shapes were defined by sequences of 2D views (frames) of dots on a rotating 3D surface. (1) Subjects' accuracy of shape identifications dropped from over 90% to less than 10% when either the polarity of the stimulus dots was alternated from light-on-gray to dark-on-gray on successive frames or when neutral gray interframe intervals were interposed. Both manipulations interfere with motion extraction by spatio-temporal (Fourier) and gradient *first-order* detectors. *Second-order* (non-Fourier) detectors that use full-wave rectification are unaffected by alternating-polarity but disrupted by interposed gray-frames. (2) To equate the accuracy of 2AFC planar direction-of-motion discrimination in standard and polarity-alternated stimuli, standard contrast was reduced. 3D shape discrimination survived contrast reduction in standard stimuli whereas it failed completely with polarity-alternation even at full contrast. (3) When individual dots were permitted to remain in the image sequence for only two frames, performance showed little loss compared to standard displays where individual dots had an expected lifetime of 20 frames, showing that 3D shape identification does not require continuity of stimulus tokens. (4) Performance in all discrimination tasks is predicted (up to a monotone transformation) by considering the quality of first-order information (as given by a simple computation on Fourier power) and the number of locations at which motion information is required. Perceptual first-order analysis of optic flow is the primary substrate for structure-from-motion computations in random dot displays because only it offers sufficient quality of perceptual motion at a sufficient number of locations.

Introduction

A sequence of 2D projected images (frames) of a moving 3D object is sometimes perceived as a moving 3D shape. When each isolated 2D frame is uninformative about 3D shape, but the sequence causes a 3D shape to be perceived, this is called the *kinetic depth effect*, after Wallach and O'Connell (1953). When a computer algorithm recovers 3D shape from a 2D frame sequence, it is called *structure from motion* (Ullman, 1979).

There are two classes of proposed models for deriving 3D shape from 2D frame sequences; we designate them as *feature-correspondence* models and *flow-field models*.

Feature-correspondence models. Feature-correspondence models use geometric constraints, usually coupled with assumptions of rigidity, to derive shape. Examples of algorithms that derive a 3D configuration from a set of n points (or similar features) displayed in each of m frames are Hoffman & Bennett (1985) and Ullman (1979, 1985), or see Braunstein, Hoffman, Shapiro, Andersen, & Bennett (1987) for a more empirical treatment. A list of visual features is identified and located in 2D space on each frame. In this class of model, the correspondence of point n in frame m with equivalent point n in frame $m+1$ is assumed to be known. Using Euclidean geometry and the assumption of object rigidity, a 3D location for each feature on each frame is derived. The set of 3D locations determines object shape.

Flow-field models. Flow-field models derive object shape from local velocity information described by optic flow fields. An object is described by many points or other features densely scattered on its surface and possibly throughout its volume. The flow-field is computed from the velocities of groups of points over a sequence of frames. Flow-field

velocities determine relative depths and orientations and thereby object shape (e.g., Clocksin, 1980; Hoffman, 1982; Koenderink & van Doorn, 1986). Flow-field models suggest that a sequence of frames might be considered not as an abstract list of features with associated location information, but as a motion stimulus to one or more motion-detection mechanisms. In this article, we are primarily concerned with determining the nature of this motion stimulus.

First-Order and Second-Order Motion Systems

We consider here three kinds of motion-detectors: two first-order detectors, which we designate as (1) spatio-temporal motion energy detectors and (2) gradient detectors, and (3) second-order detectors. A first-order detector detects motion in stimuli that would yield motion to a local spatio-temporal Fourier analysis; a second-order detector may detect such motion but also detects motion in a wide class of stimuli that do not yield directional motion under any kind of Fourier analysis. We examine these kinds of detectors in more detail below.

Fourier motion-energy detectors: the elaborated Reichardt detector (ERD). Low-level motion mechanisms are now thought to be based on systems that approximate a local spatio-temporal Fourier analysis of frame sequences (Adelson & Bergen, 1985; van Santen & Sperling, 1985; Watson & Ahumada, 1983; Watson, Ahumada & Farrell, 1986). Indeed, whenever the spatio-temporal frequency components of a stimulus differ in temporal frequency, the output of these mechanisms is simply the sum of their responses to the individual spatio-temporal Fourier components of the stimulus (derived from their equivalence to Reichardt detectors -- van Santen & Sperling, 1984). The Reichardt detector (Reichardt, 1957) was the first computational motion detector. The elaborated Reichardt detector (van Santen & Sperling, 1984a, 1984b, 1985) successfully extended the basic scheme to the prediction of human psychophysical data, although there were earlier attempts (e.g., Foster, 1969, 1971). The

motion models of Watson & Ahumada (1983) (when elaborated) and of Adelson and Bergen (1985) have motion-detection mechanisms that are defined differently but have been shown to be equivalent to Reichardt detectors at their final outputs (van Santen & Sperling, 1985), although the order of intermediate operations is different.

Motion discrimination (e.g., the discrimination of leftward from rightward motion) now appears to be a different process than velocity discrimination. The elaborations of the basic motion-detection mechanism to account for velocity discrimination are quite complex (e.g., Watson & Ahumada, 1985; Heeger, 1987) and involve the interplay of many elementary motion detectors. Since all these models ultimately depend on a basic mechanism that is equivalent to an elaborated Reichardt detector (ERD), we shall describe the ERD in more detail.

Insert Figure 1 here.

A Reichardt motion detector consists of two component half-detectors. One half-detector compares the intensity at point A , time t with the intensity at point B , time $t + \Delta t$ (see Figure 1). The other half-detector looks at (B, t) and $(A, t + \Delta t)$. While each half-detector can detect motion by itself, the two together have some important advantages. They signal motion in opposite directions by outputs of opposite sign, and by canceling evidence for movement in opposite directions, they help to disambiguate flicker and other nonmotion stimuli from true motion.

To account for psychophysical data, the spatial points A and B are replaced with spatio-temporal receptive fields, I_A and I_B , and the pure delay Δt is replaced with a linear filter.

The receptive fields I_A and I_B determine the spatial orientation-tuning of the detector, and I_A and I_B taken with the time delay Δt jointly determine the velocity tuning. Theories of human motion perception which we have discussed assume that populations of such detectors exist in different sizes (scales) and at each scale they are tuned to different orientations and velocities. The aggregated outputs of all these detectors are combined by a *voting* (decision) rule to predict the direction of perceived motion at each spatial location and time.

ERDs (and hence the various equivalent spatio-temporal motion-energy models) account for a wide variety of critical data on direction of motion discrimination (van Santen & Sperling, 1984a, 1985). To provide velocity sensing, outputs of arrays of basic spatio-temporal motion detectors must be combined (Watson & Ahumada, 1985; Heeger, 1987), because an isolated ERD will not function adequately as a velocity detector. Stimulus contrast and many factors relating to velocity tuning are confounded in the response of any one motion detector. Watson and Ahumada (1985) propose direct coding of the temporal frequency of sets of motion detectors. Heeger (1987) compares the overall pattern of responses of a set of motion detectors to an unknown stimulus to the patterns produced by known training stimuli.

Gradient detectors. A second class of first-order motion detection mechanisms uses gradients in the computation. Examples are Limb & Murphy (1975), Fennema & Thompson (1979), Horn & Schunk (1981), Marr & Ullman (1981), and Harris (1986). Basically, these models find local areas where luminance $I(x, y, t)$ varies as a function of (x, y) , i.e., has a nonzero spatial gradient $\nabla I(x, y, t) \neq 0$. The velocity v is determined by the ratio of the change in $I(x, y, t)$ as a function of time to the change in $I(x, y, t)$ as a function of space. Gradient models do a single local computation that embraces both the Reichardt motion detection mechanism and the subsequent velocity stage of the flow-field models.

Whenever the spatial luminance gradient is small, velocity estimates are extremely unstable. Therefore, Adelson & Bergen (1986) proposed weighting the local velocity estimates by a "confidence" value. Choosing the "confidence" level as the local value of the squared gradient converts the gradient computation into a least-squares estimate of velocity (Lucas & Kanade, 1981), a computation that can be carried out by the first-order motion-energy/elaborated-Reichardt systems that we outlined above. Thus, while at first glance gradient computations seem quite different from Fourier first-order computations, the difference vanishes when a realistic gradient computation is made (Adelson & Bergen, 1986).

Second-order motion detection. Stable perception of direction of movement and of velocity can arise from complex stimuli which are essentially invisible to first-order motion detectors – they fail to report any consistent direction (Chubb & Sperling, 1988a, b). Motion detectors to perceive Chubb and Sperling's motion stimuli require two stages of linear filtering separated by a full-wave rectification stage that computes the absolute value of contrast. For the present stimuli, however, the linear filtering stages are unnecessary and will be omitted. Because of the necessity of a two-stage analysis (first rectification with or without filtering, then Reichardt-or-equivalent motion detection), motion detectors that can detect such stimuli are called *second-order*. Early evidence (Chubb & Sperling, 1987) suggests that second-order systems may operate primarily foveally and with lower spatial resolution than first-order detectors. Since they depend on rectification, with inevitable loss of information, second-order systems have higher contrast thresholds than first-order systems (Chubb & Sperling, 1989a,b).

First-order and second-order systems and KDE. This paper asks whether the *ability* of humans to perceive 3D shape from a 2D frame sequence depends on the strength of evidence supplied to first-order motion mechanisms. This question stands in sharp contrast to much of the historic work on kinetic depth effect, which emphasized cues such as perspective

(i.e., Braunstein, 1962), numerosity (Green, 1961), or occlusion (Anderson & Braunstein, 1983) and their effect on the *nature* of a shape percept. We ask whether strong input to a first-order motion system is necessary to support shape perception. Our strategy is to introduce factors such as flicker or contrast (polarity) reversal that weaken or disrupt a first-order motion mechanism. We can then ask whether the ability to perceive 3D shape is especially degraded. Symmetrically we ask, do second-order systems support 3D shape perception?

Insert Figure 2 here.

In the experiments of this paper, kinetic depth displays are rendered as dots scattered randomly on a 3D surface. These are projected as a 2D image of bright dots on a neutral gray background. Figure 2a schematically illustrates a spatio-temporal analysis of a moving intensified (brighter) dot on a gray background. A frame sequence defines the stimulus as a function in (x, y, t) , where x and y represent locations in the picture plane, and t represents frames (time). Figure 2 simplifies the analysis by showing only the (x, t) plane. A line in the (x, t) plane represents the x -component of velocity. A spatio-temporal receptive field here tuned to precisely the velocity of the illustrated points is a core component of one representational form of the Fourier energy motion detectors (Adelson & Bergen, 1985; Watson & Ahumada, 1984; and by equivalence, the ERD, van Santen & Sperling, 1984).

Figure 2b illustrates a manipulation which intersperses gray frames between motion samples, but maintains the same velocity. This reduces the amplitude of the fundamental motion component by half and introduces many low-amplitude motion components opposite in direction to the fundamental. One such opposite direction detector is illustrated in Figure 2b. An alternating gray frame display is equivalent to a half-wave rectification of a polarity

alternation stimulus (see below). For our gray-frame stimuli, the total Fourier energy in each direction is approximately equal. If the sensitivities to the various spatio-temporal motion components were equal, the energy in each direction would balance and neutralize the Fourier system. Empirically, at constant velocity, reducing the number of samples (as in a gray frame versus a standard motion stimulus) always impairs the perceived quality of stroboscopic motion (Sperling, 1976). Reducing blank (background level) interstimulus intervals to about 20 msec (and hence varying velocity) improves planar apparent motion between two alternating frames of random dots (Braddick, 1973, 1974) or multi-frame sequences (Burt & Sperling, 1981).

Figure 2c illustrates a motion stimulus which alternates polarity of the motion token between intensities higher and lower than the neutral (mean) gray level. Polarity alternation provides cancelling inputs to local spatio-temporal filters tuned to the "veridical" motion direction; alternation, as illustrated, stimulates large-scale detectors tuned to the opposite direction (Anstis, 1970; Anstis & Rogers, 1975; Chubb & Sperling, 1988b; Rogers & Anstis, 1975). Like the spatio-temporal energy models, the gradient methods, which examine changes in luminance patterns over time, are also disrupted by polarity reversal.

We investigate interspersed gray frames and polarity reversal (and other manipulations, see Landy, Sperling, Doshier, & Perkins, 1988) that may disrupt first-order processes. We determine whether 3D shape extraction is disrupted. It is also important to determine whether any such disruption is special to 3D shape extraction processes, or whether it can be accounted for exactly by decrements in simpler 2D visibility and motion tasks.

The objective measure of 3D shape recovery. The essence of kinetic depth perception is the addition of depth information to a 2D image to create a perception of a 3D object shape. We ask whether kinetic depth percepts depend on first-order motion analysis. In order to have

more than a qualitative answer to this question, it was first necessary to develop an objective index of 3D shape perception. To this end, we (Sperling, Landy, Doshier, & Perkins, 1987) developed a shape identification task with a very low guessing baserate (near 2%) and a large performance range (up to 95+%). This task requires subjects to identify a display as depicting one of a large lexicon (53) of three-dimensional (3D) surface shapes. In this paper, we also use comparison tasks such as detection, direction discrimination and motion segmentation in several control studies (Footnote 1).

General Methods

Apparatus. Stimuli were pre-generated and stored on a Vax 11/750 computer that shipped images to an Adage RDS-3000 image display system. A Conrac 7211C19 RGB color monitor was used for display, operating at a refresh rate of 60 Hz, noninterlaced. Only the green beam of the monitor was used.

Procedure. Displays were seen through a viewing tunnel and circular aperture, which provided monocular viewing at a viewing distance of 1.6 m. The circular aperture was slightly larger than the displays. The size, intensity, timing and content of the displayed frame sequences are listed below for each experiment separately. Following each display sequence, the subject pressed keys or typed the required judgment. The primary task was shape identification. Control tasks included standard two-interval detection, direction-of-motion discrimination, and motion segmentation. Displays were viewed in mixed lists within experiments.

The methods sections for Experiment 1 to Experiment 6 are presented together below, in the order in which the results will be discussed. This allows an uninterrupted presentation of the arguments in the Results section, where motivation for the particular conditions and

experiments can be found. The experiments were actually run in the following order: 1, 3, 5, 2, 6, then 4.

The displays, or conditions, for Experiments 1, 2, and 3 – the 3D shape identification experiments – are summarized in Table 1. The displays, or conditions, for Experiments 4, 5, and 6 – planar motion experiments – are summarized in Table 2. Distinct display types are numbered continuously in the two Tables.

Method: Experiment 1 (Main)

Identification stimuli. The main experiment compared objective performance levels on standard kinetic depth displays with performance on comparable displays that disturb or weaken first-order motion cues. The objective measure was percent correct identification. The shape lexicon was based on peaks, valleys, and flat regions located in one of two triangular layouts. Figure 3a shows the two triangular layouts on a square ground, and Figure 3b shows some examples of shapes. Figure 3c illustrates a shape movement, and Figure 3d indicates the size of a single display frame. Stimulus identification consisted of reporting the layout (Up vs Down), the sign of the bump (+ = peak, 0 = flat, - = valley) in each of locations 1, 2, and 3, and the direction of rotation. (See Sperling, Landy, Doshier, & Perkins, 1987, for details.)

For the 3D shape identification task, feedback consisted of a list of the correct responses. For any stimulus, there were two correct responses, which are depth-reversals of one another; the depth reversals are coupled with opposite perceived directions of rotation. Subjects were initially shown perspective drawings of shapes and instructed in naming performance. Subjects were trained in practice sessions until they achieved approximately 85% correct on the easiest

stimuli.

Insert Figure 3 here.

The standard kinetic depth display consisted of white dots on a mid-intensity (gray) background. The displays were 300 dot random subsamples of the picture plane, displayed with an x,y resolution of 182×182 pixels (Footnote 2). Projections were parallel. Peaks or valleys had simulated height equal to half the side of the square ground. The smooth surface was constructed by smoothing of a spline interpolation over the stimulus peaks and the ground. The surface was initially parallel to the projection plane, and rotated first right (or left) 25 deg, back through to left (or right) by 25 deg, and then back full-forward (25 deg amplitude sinusoidal rotation) over a period of 30 new image frames. Stimulus edges never appeared in the display window. The displays assumed no occlusion of dots by the 3D surface (transparency). The standard display rate was 15 new frames (with changed frame contents) per second. Each new frame was shown for 4 sync cycles, at a monitor sync rate of 60 Hz. Half speed displays either showed new frames every 8 sync cycles, or at 4 sync cycles with interleaved gray frames. In the data of Sperling *et al.* (1987), a similar white-on-black display condition yielded identification performance in the 95% range. Other conditions modified this standard display.

Display geometry and timing. The 3D shape display was confined to the central 182×182 pixels of a 512×512 raster (60 Hz, no interlace). Background luminance was uniform over the entire 512×512 area. The 182×182 display area subtended 3.7 by 4.2 deg at a viewing distance of 1.6 m that was controlled by viewing tube and aperture. On each trial, a fixation spot appeared for 1 sec, followed by 1 sec of blank (gray) screen, then the rotating

stimulus for 2 sec (4 sec for half-speed displays). The screen was blank until the next trial was initiated. Responses were typed into a separate keyboard, and feedback (correct stimulus identification) appeared on a separate CRT.

Calibrated intensities. The display monitor was calibrated to equate the light and dark dots on the gray background, i.e., the luminance energy gain of increments and the luminance energy loss of decrements (Footnote 3). Three subjects participated. For subject MSL, the standard intensity condition consisted of background luminance of 31.8 cd/m^2 (average of $11.6 \mu\text{cd/pixel}$) with $13.2 \mu\text{cd}$ additional (or lowered) intensification for each stimulus *dot* (at viewing distance of 1.6 m). For subject CFS, the background was 31.0 cd/m^2 (average of $11.3 \mu\text{cd/pixel}$) with increments or decrements of $13.2 \mu\text{cd/dot}$. For subject JBL, the background was 38.8 cd/m^2 (average of $14.2 \mu\text{cd/pixel}$) with increments or decrements of $20.9 \mu\text{cd per dot}$. Note: Subject CFS could not be refracted completely to normal vision; his corrected Snellen acuity was approximately 20/40. All other subjects had normal or corrected-to-normal vision.

Conditions. The main experiment included 11 display conditions. Each of the 54 possible shape stimuli appeared once in each of the 11 conditions, for 594 identification trials per subject. All of these stimuli were shown in one large mixed list, divided over 4 sessions.

Insert Table 1 here.

The relevant characteristics of the 11 display conditions are listed in Table 1. All displays in this experiment, except condition (11), depict the motion of 3D shapes in 2D projection. An unconstrained subsampling of points on the 3D shapes, includes density cues that result when peaks and valleys cause dots to bunch together in the projection of the 3D

surface onto the 2D image plane. Except in displays in conditions (1), (3), and (13), subsampling of dots was constrained such that local density was constant across the display. Density cues were eliminated from the image sequences by adding or subtracting a small number of points on each frame so as to equate dot density within local regions comprising approximately $1/10 \times 1/10$ th of the stimulus area. Constant-density subsampling introduced minor levels of apparent scintillation. The amount of scintillation can be expressed as the average percentage of dots *not* maintained from frame m to frame $m+1$, or equivalently, in the expected lifetime of dots. Over all the density-controlled (no density cue) displays in the experiment, the average scintillation was 5%, yielding an expected dot lifetime of 20 frames for the dots of frame 1. (These displays are indicated as ≤ 30 in Table 1.) Condition (11) extracts the local density cues in (1), but eliminates systematic motion information. Time- and position-dependent density is generated by random sampling from the rotating 3D shape with a new random sample for each frame (dot lifetime of 1 frame). This destroys systematic motion cues, but maintains local variations in dot density under rotation.

Most displays depict a standard rotation speed as described above. In conditions 3 and 4, half-speed rotation is produced by displaying each new frame for 8 (sync) repetitions (instead of 4 in the standard condition). The half-speed gray frame condition (7) is accomplished by interleaving 4 repetitions of each new frame with 4 repetitions of gray frame. Full-speed gray frame condition (8) is accomplished by interleaving 4 repetitions of *every other* new frame of the standard stimulus with 4 repetitions of gray frame.

Standard displays depict the 3D shapes by displaying bright dots, of a selected standard intensity of increment on a neutral (gray) background. Intensity listings in Table 1 refer to a multiple of the standard dot intensification, positive for increments and negative for decrements. In alternating polarity displays, the dots are bright in odd frames, and dark on

even frames (labelled 1 : -1). In alternating gray displays, gray background is displayed on all even frames (labelled 1 : 0). Other nonstandard increments serve as controls.

Method: Experiment 2 (Equated Contrast Identification)

Conditions. The task in this experiment was 3D shape identification; it was conducted with displays that had been equated for discrimination of motion direction by reducing dot intensity by an amount determined from Experiment 5. Subjects viewed *standard* 3D shape identification displays – Table 1, condition (2) – in which the dot increments had been reduced (condition 12). The data for the standard (non-alternating condition) in Experiment 5, by interpolation, allowed the selection of an increment intensity which would approximately equate the percent correct motion direction judgment of the standard condition with polarity alternation stimuli at full intensity increments and decrements. This equal-direction-discrimination value was determined separately for each of the two subjects. Each of the 54 identification stimuli was presented in random order.

Display geometry and calibrated intensities. Viewing conditions were the same as those described in Method Experiment 1. Calibrated intensities were: For MSL, the background intensity was 31.0 cd/m^2 with increment/decrement intensity of $8.8 \text{ } \mu\text{cd/dot}$. For JBL, the background intensity was 38.0 cd/m^2 and increment intensity was $9.6 \text{ } \mu\text{cd/dot}$.

Method: Experiment 3 (Lifetimes)

Conditions. This experiment compared three conditions in which the lifetimes of the dots were 2 frames, 3 frames and ≤ 30 frames (continuous) (conditions 14, 13, and 2, respectively, under Experiment 3 in Table 1). See Figure 6a for an illustration. New dots were subsampled randomly, with additional subsampling to eliminate density cues for all

conditions of this experiment. The task was 3D shape identification. Each of the 54 shapes appeared once in each condition, for 162 identification responses per subject.

In the 2-frame displays, each subsampled dot appears for exactly 2 consecutive new frames. Half of the dots are replaced with another random subsample on each new frame. This introduces 50% scintillation (density control does not require additional subsampling). In the 3-frame displays, each dot appears for exactly 3 consecutive new frames. One third of the dots are replaced with another random subsample on each new frame, for 33% scintillation. In the ≤ 30 -frame displays, each dot remains visible for all 30 new frames of the display, with exceptions to eliminate the density cues, which introduced 5% scintillation.. This is identical to condition (2) of Experiment 1.

Display geometry and calibrated intensities. The identification stimuli, subjects, and viewing conditions are identical to those listed in Method Experiment 1. Calibrated intensities were identical to those in that experiment.

Method: Experiment 4 (Visibility)

Insert Table 2 here.

Conditions. Conditions for Experiments 4, 5, and 6 are listed in Table 2. This experiment required subjects to detect the presence of uniform planar motion in a two-interval forced-choice (2IFC) paradigm (Figure 7a). The subject indicated which interval contained the moving stimulus. Guessing baserate is 50%. Stimuli consisted either of normal light dots on a gray background (conditions 15-19), or polarity alternating light and dark dots on the background (conditions 20-24). The five conditions of each type are measures of motion-

direction at five levels of the "standard" (condition 2, Experiment 1) dot intensity (increments or decrements). For MSL, the intensity conditions were 17%, 25%, 33%, 42%, and 50% of standard. For JBL, the intensity conditions were 33%, 50%, 67%, 83%, and 100% of standard. The ten conditions each were tested 20 times per block in random order, for 5 blocks, or a total of 1000 trials per subject.

Insert Table 2 here.

Display geometry and calibrated intensities. Each interval of the display consisted of a 1/3 sec fixation spot, 1/3 sec blank screen, followed by 1.067 sec (16 frames at 15 frames/sec) of stimulus. Non-motion intervals displayed uniform gray fields. Motion intervals displayed a sequence of approximately 17 random dots in a 48×48 pixel (.97 by 1.1 deg) patch (0.0075 dots/pixel, or 16 dots/deg² average density) moving left or right by 1 pixel/frame, or approximately 0.35 deg/sec. The viewing conditions were identical to those described above for Experiment 1. For MSL, the background intensity was 31.0 cd/m², and the intensity increment or decrement was 13.2 μ cd/dot at 100% standard intensity. For JBL, the background was 32.0 cd/m² and the increment or decrement was 16.9 μ cd/dot at 100% standard intensity.

Method: Experiment 5 (Motion Direction)

Conditions. The task in this experiment was discrimination of leftward from rightward motion of dots within a square in the center of a larger field (Figure 8a). The stimuli were a uniform field of dots of the approximately the same density as the shape identification stimuli of Experiments 1-3. The drift speed of dots in the central square (0.35 deg/sec) was

approximately the average of ground dots at the edges of the shape identification stimulus, or approximately one eighth of the peak velocity in that stimulus. In the 3D shape identification stimuli, peak speed is achieved for only one or two frames and then only at the exact center of a peak or valley. Most dots in the vicinity of a peak or valley have an average speed of one-half peak speed or less. The selection of drift speed for this direction of motion control is considered in the Results section.

The dots were either all white on gray (standard) (conditions 25-29) or alternated in polarity (conditions 30-34) from frame to frame. Standard and alternating images were crossed with five increment intensity levels at 33%, 50%, 67%, 83%, and 100% of the standard increment/decrement. Each of the ten conditions had 200 samples, 100 with each movement direction, for a total of 1000 direction judgments per subject.

Display geometry and calibrated intensities. Each trial consisted of a 1/3 sec cue spot, 1/3 sec blank gray frame, and 1 sec motion display, followed by a blank frame during the response interval. The image was 200×200 pixels, 4.1 by 4.6 deg at a viewing distance of 1.6 m. This included a dynamic noise background, with a moving center of 48×48 pixels. Dot density was approximately 16 dots/deg², and drift velocity was 1 pixel/frame, or approximately 2.3 arcmin/frame, or 0.35 deg/sec. The viewing conditions and calibrated *standard* intensities are the same as those in Method Experiment 1.

Method: Experiment 6 (Motion Segmentation)

Conditions. The task in this experiment was motion segmentation. Each display consisted of a 3×3 grid of patches of planar motion, with eight patches drifting left (in a left-drifting surround) and one patch drifting right, or vice versa (Figure 9a). The subject's task was to name the location and direction of the odd motion.

There were two conditions in this experiment: bright dots of standard intensity (35), and dots of alternating polarity (36) on a gray ground.

For JBL, all conditions were intermixed, such that each of three blocks showed 72 stimuli from condition (35), and 54 stimuli from each of condition (36) and a third condition which we do not report here. For MSL, two blocks had 90 trials each of conditions (35) and (36).

Display geometry and calibrated intensities. Each image was 200×200 pixels or 4.05 by 4.62 deg at a viewing distance of 1.6 m. Each motion patch was 48×48 pixels filled with dots at a density of approximately 17 dots/patch ($.0075$ dots/pixel, or 16 dots/deg²), of 1 pixel/frame drift. The background moved in the same direction as the common-motion patches; the odd-motion patch moved in the opposite direction. Other viewing conditions were the same as in previous experiments. For MSL, background intensity was 31.0 cd/m², with increment/decrement intensity of 13.2 μ cd/dot for conditions (1) and (2). For JBL, background intensity was 38.0 cd/m², with increment/decrement intensity of 19.2 μ cd/dot in conditions (1) and (2), and 9.6 μ cd/dot for the equated condition (3).

Results

Shape Identification

Elimination of the density cue (Experiment 1). When a surface is depicted by a random sampling of surface points which then undergo rotation, local regions of higher or lower dot density change over rotation. To assess the possibility that these changes in dot density *per se* can be used as cues to 3D shape, identification performance for image sequences that include both motion and density cues is compared to those in which density cues are eliminated, or in which only the density (but not the motion cues) are preserved. (See Method

Experiment 1 for experimental details.) Relevant individual subject data are shown in Figure 4. (These results were initially reported in Sperling, Landy, Doshier & Perkins, 1989). Eliminating density cues from motion sequences has only a small effect on the subjects' ability to identify shape from strong structure-from-motion stimuli, which may actually be due to introduction of scintillation. One of the three subjects (MSL) was able to perform significantly above the 1.9% guessing baserate (29.6%) with density cues alone in the absence of motion cues, by using a sophisticated guessing strategy. Since our conditions involve the disruption of strong input to low level motion systems, it was desirable to eliminate any cue, such as density, which might contaminate estimates of shape identification with weak structure from motion image sequences. Therefore, all other displays exclude the density cue. All critical image sequences were constructed to have uniform dot density in local regions of the image plane.

Insert Figure 4 here.

Insert Figure 5 here.

Standard sequence: Motion without density cue, standard and half-speed (Experiment 1). Percent correct 3D shape identification is shown in Figure 5. Standard errors of all proportions in the Figure are less than 6%; chance is 1.9%. The 3D shape task is illustrated in Figure 3. *Standard* sequence conditions display sampled dots which are a fixed increment brighter than the gray background. Percent identification levels are shown for "standard" rotation speed (sinusoidal rotation of amplitude 25 deg and period 30 frames, at

frame rate of 15 new frames/second), and for half speed (7.5 new frames/second). The average percent correct is similar for both speeds, with half-speed slightly less for subjects JBL and CFS.

Gray frame dilution (Experiment 1). By interspersing a background level (gray) blank frame between each frame depicting points of the object, we presented direction-ambiguous information to first-order motion mechanisms while maintaining the visibility of the dot features in any given frame (see Discussion section: Fourier Analysis of the Stimuli). There were two variants of this manipulation: one which equated the viewing time for each new image seen, but consequently slowing the rotation rate of the stimulus in time; and the other which replaced every other new stimulus frame with a blank frame, but equated effective rotation rate. Both of these variants destroyed the ability to recover shape information from the stimulus (see Figure 5). Only one of three subjects (MSL) maintained significantly above chance performance (average of 11%) on image sequences with alternating gray frames. Although this represents above chance identification performance, it is dramatically worse than his identification performance of nearly 90% with the unperturbed standard sequence. Rotation speed in these ranges had only small effects on either standard or alternating-gray conditions, and thus can not account for the impact of alternating gray frames on 3D shape performance.

Alternating polarity (Experiment 1). In polarity alternation, the stimulus tokens (subsampling dots on the shape surface) alternate between intensity increments and decrements (light on gray then dark on gray) on each frame. Adjacent image frames primarily support motion signals of the incorrect sign in the first-order system. Analysis of the change in location of these motion signals over many frames, or analysis following some form of rectification (second order, or non-Fourier analysis, see Chubb & Sperling, 1988b) could support the correct motion interpretation. Two levels of polarity alternation were examined,

one with light dots equal in intensity to those in normal image sequences and one with light dots half the intensity of those in normal image sequences. In both cases, the dark dots were symmetrically below the background level. Again, disrupting the input to low level motion systems reduced shape identification performance to near guessing baserates. Only one of three subjects (MSL) retained above-chance identification on polarity alternation stimuli (average of 10%).

Intensity alternation stimuli (Experiment 1). Introducing blank (gray) frames between every stimulus frame in an image sequence causes ambiguous signals in the first-order motion systems. Introducing polarity reversal caused direction-reversed signals in the first-order motion systems. Both manipulations also introduce whole-screen flicker; stimulus frames including intensified dots appear every other new frame for a flicker frequency of 7.5 Hz. We included several *contrast alternation* (without polarity alternation) conditions, which also exhibit whole-screen flicker at 7.5 Hz, both of which sustain performance levels close to that of the standard stimulus.

One flicker control alternated the intensity of stimulus points between the intensity level in normal displays and twice that. This stimulus is the sum of the standard stimulus and the gray frame stimulus. The other flicker control alternated between 1.5 and .5 the standard levels. This stimulus is the sum of a half contrast standard and the full-contrast gray frame stimulus. Alternatively, this stimulus can be decomposed into a standard stimulus plus a half-contrast polarity alternation stimulus (i.e., a high-flicker added stimulus). The performance levels on both control conditions are quite consistent with a Fourier power (first-order) analysis of these sequences (see the Discussion). Thus, addition of flicker *per se* does not account for the decrements in performance for alternating-gray and alternating-polarity displays.

Equated intensity control (Experiment 2). We have demonstrated that gray frame alternation and polarity alternation both severely disrupt the ability of subjects to extract 3D shape from an image sequence which allows highly accurate 3D shape identification under normal display conditions. However, perhaps this disruption is not unique to the recovery of depth information. Perhaps it simply reflects a general disruption in visibility or motion discrimination. In order to control for this possibility, we constructed equated-intensity controls based on performance in simple direction-of-motion discrimination. The details of the direction discrimination data are described below and in the Method for Experiment 5. By reducing the intensity (lowering contrast and hence visibility) of a normal (light on gray background) planar motion stimulus, it is possible to make it equivalent to a full-intensity polarity alternation stimulus for the purposes of left-right direction-discrimination. The direction-discrimination displays present a patch of moving dots of approximately the same area as a bump in the 3D shape displays. Having found the equivalent reduced-contrast normal stimulus, we then compared 3D shape discrimination for the two stimuli (reduced-contrast normal, full-contrast polarity alternation). These results are shown on the extreme right in Figure 5 for MSL and JBL. If the effect of polarity alternation can be attributed solely to a visibility-related decrement, then the equivalent intensity condition should have yielded equal shape identification performance to that for polarity alternation. In fact, lowering intensity adversely affected shape identification, but levels were still well above those for shape identification from polarity alternation displays. The percent identification for normal, equivalent intensity and polarity alternation conditions were 87%, 43%, and 15%, respectively, for MSL, and 69%, 33%, and 6%, respectively, for JBL. (Standard error of the 43% and 33% equated contrast conditions is $\pm 6\%$.)

Tracking disruption – Lifetimes (Experiment 3). We have shown that conditions which disrupt input to low-level motion analyzers also eliminate the ability to perceive three-dimensional shape, at least in the conditions of our experiments. It is interesting to contrast this with a manipulation which eliminates the ability to track individual image features (dots) over multiple frames. Models that emphasize the extraction of specific image features and their image plane location (Hoffman & Bennett, 1985; Ullman, 1979, 1985, etc.) might predict that eliminating feature stability should have an equally large impact on the shape identification. We investigated this hypothesis by comparing feature stability over a full 30 frame image sequence with stimuli in which features (surface dots) were stable for only 3 and 2 frames, after which they were replaced with a different random sample of dots (Figure 6a). The shape identification data are shown in Figure 6b. For two subjects (MSL, CFS), reducing tracking to two frames (and increasing scintillation substantially) had very little effect on performance. A third subject's (JBL) two-frame lifetime identification performance was about 54% of normal. While this was a 2 \times loss, it was a much smaller loss than the 10 \times loss induced by polarity alternation for JBL. Thus, feature-tracking models of the kinetic depth effect appear unable to account for the performance in our experiments.

Insert Figure 6 here.

Motion Visibility, Discrimination and Segmentation

This section compares the disruptive effects of polarity alternation on 3D structure-from-motion (shape identification) to its effects on visibility, direction-of-motion discrimination and segmentation.

Motion visibility (Experiment 4). Subjects were asked to detect which of two temporal intervals contained a motion stimulus and which contained a uniform field of background intensity. The motion stimulus was either a random dot field moving at uniform velocity to the right or left, or a polarity alternation version of the same stimulus. The display is schematically illustrated in Figure 7a. The size of the region was approximately that of a single peak or valley in the shape displays, of approximately the same dot density and a representative velocity (between that of the ground and maximal velocity of a peak or valley). (See Method Experiment 4 for details.) Detection may reflect contributions by non-motion systems. For example, Watson & Ahumada (1985) claim that detection of moving stimuli with velocity less than 2 deg/sec is performed by non-motion systems.

The detection data are shown in Figure 7b. Across a range of stimulus intensity increments (17%-50% of normal level intensity for MSL, 33%-100% of normal level intensity for JBL), the effect of polarity alternation was small. For MSL, normal and polarity alternation displays (averaged across contrasts) yielded 73% and 74% correct detection, respectively. For JBL, the figures were 83% and 90%, respectively. Whereas polarity alternation almost destroys the ability to extract three-dimensional shape, it may slightly improve stimulus detection relative to standard displays for our conditions. Detection accuracy with polarity alternation is essentially perfect at intensity levels comparable to those used in the 3D shape experiment (MSL at 50% intensity is 95% correct, and JBL at 100% intensity is 96% correct).

Insert Figure 7 here.

The small effect of polarity alternation on detection performance is consistent with the near-symmetry of increment and decrement thresholds in small-target pedestal detection experiments (Krauskopf, 1980; Rashbass, 1970; Roufs, 1974), although some studies find decrements slightly easier to detect (Patel & Jones, 1968; Short, 1966). Alternatively, the fundamental flicker component of the polarity alternation stimulus is 7.5 Hz, approximately at the peak of the flicker sensitivity function (Watson, 1986), which suggests that polarity alternation may be most sensitively detected by flicker-sensitive mechanisms.

Direction-of-motion discrimination (Experiment 5). Subjects were asked to discriminate the direction of motion (right or left) of a small patch of random dots moving with uniform velocity (Figure 8a). The dots were either always light against the background, or alternated polarity from frame to frame. Discrimination was examined over a range of intensity increments (or decrements) per dot. (See Method Experiment 5 for details.)

Direction discrimination data are shown for two subjects in Figure 8b. Polarity alternation impaired subjects' ability to discriminate motion direction: Averaged over intensity level, normal and polarity alternation conditions yielded 85% and 69% correct, respectively, for subject MSL, and 90% and 67% respectively for JBL. However, at the intensity levels that were investigated in the shape identification experiments, levels of direction discrimination for polarity alternation stimuli were good: 87% correct for MSL and 88% correct for JBL. Intensity-based decrements for standard displays in this experiment were used to select the "equated intensity" condition listed above for shape identification.

Insert Figure 8 here.

The patch size in the direction-of-motion displays was selected to be approximately the size of a bump or depression in the shape displays. The speed of drift (0.35 deg/sec) was selected to be representative of the modest speeds in many points of the 3D shape displays, where peak speeds may range up to 2.5 deg/sec. Based on data from direction of motion discrimination in near-threshold sine wave stimuli (Ball & Sekuler, 1979; Burr & Ross, 1982; Green, 1983; Watson, Thompson, Murphy, & Nachmias, 1980) and theoretical computations on direction of motion discrimination for random dot stimuli (Nakayama, 1985; Van Doorn & Koenderink, 1982), we picked the weakest motion stimulus that could be derived from the 3D shape task: the slowest reasonable speed and approximately the same number of dots in the displays to be comparable. That the direction of motion of this stimulus is nearly always judged correctly at standard intensities implies that direction of motion at a *single* location is almost completely intact when 3D shape identification is at zero.

In two-frame experiments or multi-frame experiments where two frames appear alternately, polarity alternation may lead to below chance performance on direction discrimination (Anstis, 1970). Polarity alternation excites first-order (Fourier) spatio-temporal sensors for motion opposite to the veridical direction, as schematically illustrated in Figure 2c. Here, in multi-frame movement, the (temporally and spatially) local support for movement in the opposite direction is apparently more than offset by second-order (non-Fourier) processes sufficiently often that direction discrimination rarely falls below 50%. Chubb & Sperling (1988a, 1989a,b) show that the relative dominance of the first-order and second-order information in polarity alternation stimuli depends on the spatial scale (near viewing distances favor second-order information).

Motion segmentation (Experiment 6). In contrast with simple detection or discrimination of motion direction, a more complex motion direction task did show decrements

in performance more comparable to those seen in shape identification. We developed a motion segregation paradigm in which nine small patches of uniformly moving dots were presented as a 3×3 grid embedded in a border of moving random dots (Figure 9a). All but one patch depicted motion in the same direction (left or right), while the odd patch depicted motion in the opposite direction. The stimulus dots either remained above the background level (light on gray), or alternated polarity. (See Method Experiment 6 for details.) In this situation, polarity alternation had a large impact on selection of the odd patch. MSL reported 95% correct locations with the normal display, but only 22.2% with polarity alternation. JBL reported 84% correct and 10.5% respectively (chance = 11.1%) (Figure 9b). The accuracy levels for polarity alternation displays are consistent with sophisticated guessing (see Discussion).

Insert Figure 9 here.

Discussion

Fourier and Non-Fourier inputs to structure from motion. Vivid 3D shape percepts which allow accurate 3D shape identification can arise from appropriately constructed 2D image sequences depicting projections of those shapes under rotational motion. Typically these 2D sequences provide good input to first-order spatio-temporal ("Fourier") motion analyzers. In order to determine whether strong Fourier motion is a prerequisite to shape extraction, we examined display manipulations which maintain the identity-correspondence between points in successive frames, but disrupt first-order analysis. Interleaving blank frames or alternating token contrast-polarity both had devastating consequences for the ability to identify 3D shape in our displays. The inability to recover shape was not due to overall display flicker since same-sign alteration in the intensity levels of particular tokens did not seriously disrupt

performance. Subjectively, a sensation of local motion was maintained, and selected points could still be tracked. Nonetheless, this information was not adequate to support shape identification.

The dependence of 3D shape perception on unambiguous first-order (Fourier) motion inputs suggests that, for our stimuli, direction and velocity serve as the primary input to a subsequent shape-extraction (structure from motion computation, e.g., Koenderink & van Doorn, 1986). Obviously the velocity information must be computed simultaneously or nearly simultaneously at several locations in order to perform the 3D shape task.

The main alternatives to local velocity-based computations depend on geometric analyses of identified feature elements and operate over more than two frames (e.g., Ullman, 1986). These alternative schemes are challenged by our finding that shape extraction is little affected by change in feature elements as often as every two frames. Further, subsequent work (Landy, Doshier, Sperling, & Perkins, 1988) shows that motion displays of only two-frames also support moderately good shape identification.

Williams and Phillips (1986, 1987) report what they consider a surprising perceptual phenomenon of perceiving a 3D shape in a random-dot flow field. We interpret their finding here as further evidence that a local velocity computation is the basis of perception of 3D shape. In their dynamic 2D displays, dots execute a random walk of constant step size, with displacement angle chosen from a uniform distribution with a range less than 150 deg. Subjects perceive a rotating and translating 3D cylinder. In these stochastic displays, velocity information is very similar to the local velocity information in a cylinder with dots sprinkled through its volume, rotating rigidly and translating along its axis of rotation (e.g., as displayed by Doshier, Landy, & Sperling, 1989) (Footnote 4). As in our experiments, the momentary

distribution of velocities, not the stochastic trajectories of individual dots, determines the 3D percept.

3D shape extraction is especially impaired in displays that have contradictory or ambiguous first-order (Fourier) information. Control experiments demonstrated that contrast-polarity alternation, which essentially eliminated 3D shape identification, nonetheless left the detection judgment and the direction-of-motion judgment for a small isolated moving patch quite high. Motion segmentation, which requires analysis of motion direction in a number of local display regions, was also profoundly affected by polarity alternation.

A Fourier Computation for the Strength of First-Order Motion Perception. Up to this point, we have talked in generalities about Fourier and non-Fourier computations of motion direction. Here we propose some very simple, specific, Fourier computations that account quite well for the results that we have attributed to first-order motion processes. The computation proceeds as follows.

- (1) Compute the Fourier transform of the stimulus as it was viewed by the observer, i.e., with the correct visual angles and an accurate description of the display that was actually produced. Compute the power $p(\omega_x, \omega_t)$ of each spatio-temporal frequency component.
- (2) Retain only the power p_ϵ that exceeds a small threshold $\epsilon > 0$, i.e.,

$$p_\epsilon(\omega_x, \omega_t) = \max[p(\omega_x, \omega_t) - \epsilon, 0].$$
- (3) Retain only the Fourier components that fall within a window of visibility (Watson, Ahumada, & Farrell, 1986) that includes all spatial frequencies greater than zero and less than or equal to 30 cycles per degree of visual angle and all temporal frequencies greater than zero and less than or equal to 30 Hz, vis., $(0 < |\omega_x|, |\omega_t| \leq 30)$.
- (4) The *net directional power*, DP , of all frequencies within the window of visibility is the

rightward power minus the leftward power:

$$DP = \sum_{\substack{\omega_x \cdot \omega_t < 0 \\ |\omega_x|, |\omega_t| \leq 30}} p_e(\omega_x, \omega_t) - \sum_{\substack{\omega_x \cdot \omega_t > 0 \\ |\omega_x|, |\omega_t| \leq 30}} p_e(\omega_x, \omega_t) .$$

The computation gives equal weight to all motion components within the window of visibility and zero weight to all components outside the window. In a more refined analysis, it might be useful to weight spatial frequencies according to a contrast sensitivity function. However, it is not obvious how to weight signals that are above threshold. For practical purposes, it turns out that the exact size of the window of visibility has little influence on relative *DP*s for the stimuli considered here.

Basically, the left-minus-right-difference, summed over all frequencies, is similar to the computation that is carried out by previously proposed first-order motion models. For example, within its window, an elaborated Reichardt motion detector (van Santen & Sperling, 1984) computes the algebraic sum of all velocity inputs that differ in temporal frequency. Velocity inputs that have the same temporal frequency (and therefore differ only in spatial frequency) are processed by detectors of different scales, sensitive to different spatial frequencies. Outputs of different detectors are combined at the next higher level (e.g., Adelson & Bergen, 1986).

A real detector, localized in space and time, cannot have the perfect resolution of a Fourier analysis of the entire x, y, t stimulus. The entire Fourier analysis is most appropriate for analyzing local areas where movement can be regarded as uniform and homogeneous. Even with all these qualifications, the straightforward Fourier analysis of the dot movement-

patterns is quite informative.

 Insert Figure 10 here.

Fourier analysis of the stimuli. The space-time (x,t) representations of a single dot element in each of the motion stimuli for our main conditions is shown in the left hand panels of Figure 10. The Fourier power spectra for those stimuli are shown in the right hand panels of Figure 10. Figure 10a represents a dot moving from left to right over frames. The dot is the standard intensity on the neutral background. The abscissa represents 1.07 deg of spatial position x from left to right; the ordinate represents a 1.07 sec interval of time, t , from bottom to top. The representation assumes a sampling density of 120 samples per degree of visual angle and 120 samples per sec to yield temporal discrimination up to 60 Hz and spatial discrimination up to 60 cycles per degree of visual angle. (In this representation, the four refreshes of each new image frame are seen as four repeats at the same location in alternate 1/120 sec samples. The illuminated dots on our display are depicted as 2 adjacent spatial samples.) The steep space-time function reflects the fact that our stimuli move relatively slowly (0.35 deg/sec). Figure 10b shows the corresponding Fourier power spectrum. The abscissa is ω_x and the ordinate is ω_t ; the axes cross at $\omega_x = \omega_t = 0$.

If the standard motion stimulus were moving continuously in space and time, essentially all of its components would be at the intended direction and speed. Because it is sampled in time (60 Hz refresh and 15 new frames/second) and in space (by the resolution of the pixel array) it contains ambiguous temporal and spatial components. Most of the power is in the intended direction and velocity (upper left and, symmetrically, lower right quadrants). But there is a surprising amount of power in the unintended direction as well (upper right, and

symmetrically, lower left quadrants). The ($0 < |\omega_x|, |\omega_y| \leq 30$) window of visibility is shown as the inner square in Figure 10. The computed DP strongly favors the intended direction by 5:1. Figure 10c and 10d show the stimulus representation and Fourier energy spectrum of a standard stimulus at half intensity (approximately that of the contrast-equated control). The transform is the same as 10b, but of half power. With $\epsilon=0$, the computed DP is exactly half; with $\epsilon>0$, the computed DP is less than half.

Figure 10e and 10f show the stimulus representation and spectrum for the alternating gray frame stimulus. In the case of gray-frame stimuli, power at the intended direction and velocity is halved, and approximately balanced by power dispersed over a range of velocities in the opposite direction.

Figure 10g and 10h show the stimulus representation and spectrum for the alternating-contrast polarity stimulus. In this case, the net directional power DP is of very slightly lower magnitude than for the standard stimulus, but favors the unintended over the intended direction (more power in the upper right and lower left quadrants).

Figure 10i and 10j shows the stimulus with contrast alternation between $2\times$ and $1\times$ the standard intensity. This stimulus can be viewed as the sum of the standard stimulus and the alternating-gray stimulus. Although the 2:1 contrast-alternating stimulus has some of the diffuse power of the alternating-gray stimulus, 2:1 contrast alternation puts more power into the intended direction and velocity than even the standard stimulus. Figure 10k and 10l are for stimuli with contrast alternation between $1.5\times$ and $0.5\times$ the standard intensity. This 1.5:0.5 contrast-alternating stimulus can be viewed as the sum of the half-intensity standard stimulus and the alternating-gray stimulus. The computed DP is slightly lower than for the standard

stimulus.

Tasks. The kinds of information needed for good performance in the various tasks is summarized in Figure 11 and, along with the relation to computed DP , is explained below.

Insert Figure 11 here.

Detection. In Experiment 4, we noted that simple two-interval forced choice detection (2IFC Detection) of a single local patch of moving dots is probably accomplished by other systems than the motion systems. The equality (or near equality) of detection with standard and polarity alternation displays insures that polarity alternation did not result in peripheral cancellation of the input stimulus.

Direction. Discrimination between left and right motion direction (two-alternative forced choice, 2AFC Direction) minimally requires direction (but not necessarily velocity) analysis by a motion detection system in a single location (Figure 11). As shown by the Fourier spectrum of Figure 10h, a first-order analysis of a polarity-alternation stimulus would support the unintended (opposite) direction of movement. A second-order analysis based on full-wave rectification would yield the correct direction and velocity. In full-wave rectification, the sign of contrast is lost, and the standard stimulus would be recovered. 2AFC-Direction performance is impaired by polarity alternation, but still well above chance for a wide range of contrasts. Polarity alternation leads to high levels (about 88% correct) of 2AFC-Direction performance at "standard" contrasts; hence, perceptual second-order analysis occurs under these conditions. But, alternating-contrast polarity stimuli require higher contrasts to yield equal direction-discrimination than do standard stimuli which stimulate first plus second-order

systems. This might reflect power loss in the second-order analysis, the need to overcome conflicting first-order information, or both.

Motion Segmentation. In order to isolate which of 9 patches is moving in a direction opposite to the others requires that direction of motion be assessed in several locations (Figure 11). We examine the consequences of observing (correctly perceiving the direction of motion in) n of the 9 locations. Observing just one patch, which is sufficient for the 2AFC-Direction task would lead to chance performance of one-in-nine locations – identical to the guessing level without seeing the display. Observing any two patches could improve performance by sophisticated guessing. That is, if the two patches move oppositely, then one of them is the target; if they move in the same direction, one of the remaining 7 is the target. The probability of sampling two opposite direction locations times a guessing accuracy of $1/2$ plus the probability of sampling two same directions times a guessing accuracy of $1/7$ yields an estimate of 22.2% correct. Observing any three or more patches could improve performance by a combination of informed judgments and sophisticated guessing, etc. The data for polarity alternation do not require us to consider more than two observations. Performance for polarity alternating stimuli in the odd-in-nine motion segmentation task was indistinguishable from the simple 1 in 9 baseline (11%) for one subject (10%), and slightly above the 1 in 9 baseline for another (22%), which could be achieved by sampling only two locations.

Motion segregation, like shape extraction, may be dependent on strong Fourier input largely because it requires evaluation of motion signals at more than one location nearly simultaneously. The second-order motion system operates primarily foveally (Chubb & Sperling, 1988b). Two locations might be successively fixated in our 1 sec displays. For standard displays, performance in this task is excellent (85-95%). By similar computations, this would require observation of approximately 7 locations. Thus, first-order information

supports direction of motion analysis at a number of directions simultaneously, while second-order information can support direction of motion analysis at only one or two.

3D Shape. The simplest solution to the 3D shape identification task requires simultaneous, or nearly simultaneous, knowledge of the motion-direction information (and possibly also the velocity) at the six bump locations (Sperling et al., 1989). The principle is that, to a first and adequate approximation, dots on bumps move in one direction, dots in depressions move in the opposite direction, and dots on the ground plane move very little. Thus, to solve the 3D-shape task, motion has to be categorized into 3 categories (leftward, rightward, and near zero) at a number of locations simultaneously (Footnote 5). Although the 3D-shape identification task could, in principle, be carried out with only this very coarse velocity information, more information usually is used. For example, in a version of the 3D-shape identification task with different bump heights, subjects can quickly discriminate three levels of bump height (Sperling et al., 1989). The bump-height discrimination is based on speed (Footnote 6).

Insert Figure 12 here.

Although a sophisticated local velocity computation probably underlies the 3D shape *percept*, for our set of stimuli, the simple (Fourier) net directional power, DP , computation offers an adequate account of performance in the 3D shape identification *task*. We assume that net directional power DP serves as a measure of the quality of first-order direction information in the various displays. If the 3D shape identification performance with our displays primarily depended on good first-order information, then the performance level for the various displays would increase monotonically with the quality of first-order information – here indexed by DP .

Figure 12 shows the percent correct identification in the 3D shape task as a function of computed DP for the representative 2D motion display (Figure 10a-1). DP is in units of power normalized to the standard stimulus. Identification levels increase monotonically with DP , as expected.

Full-wave rectification of polarity alternation displays (second-order processing) would allow recovery of intended motion signals. However, 3D shape identification performance on these displays is approximately at chance levels (left half of Figure 12). In principle, systematic DP favoring the unintended direction might be used in sophisticated guessing, but apparently is not. Performance on displays with polarity alternation may also reflect conflict between first-order and second-order motion information.

The effect of the power threshold ϵ in the computation of DP may be understood by comparing 3D shape performance in the contrast equated (approximately half-power standard) and 1.5:0.5 contrast alternation stimuli. Without the power threshold ϵ entering into computed DP , the contrast alternation 1.5:0.5 computed DP is only slightly higher than that for the half-intensity standard, while identification levels are quite different. However, even with $\epsilon=0$, identification performance is monotone with DP . (DP computations with $\epsilon>0$ and with $\epsilon=0$ are shown as filled and open circles, respectively, on the abscissa of Figure 12.) Hence, the 3D shape data are consistent with a DP analysis of the outputs from a first-order (Fourier) motion system.

Why First-Order Motion for 3D Shape Perception? First-order (Fourier) motion systems are assumed to be implemented with detectors like those schematized in Figure 1. Second-order (non-Fourier) motion systems may implement some form of nonlinear transformation on the image intensities prior to further spatio-temporal analysis (see Chubb &

Sperling, 1987). The two tasks in which second-order information could not be efficiently utilized, 3D shape recovery and motion segmentation, require information about motion direction (and velocity) in several local regions simultaneously. Hence, our evidence agrees with the evidence of Chubb & Sperling (1988a,b, 1989a,b) that the non-Fourier motion systems are most effective at large spatial scales, with foveal presentation, and do not function well in noncentral locations. For our stimuli, 3D structure was extracted primarily from first-order motion information.

Our stimuli were modestly complex but continuous surfaces in depth. The surfaces were depicted by randomly scattered and unconnected dots. Object transparency (where a portion of the stimulus which is behind a nearer portion of the surface can be seen) was allowed, but rarely occurred. (This form of representation is most similar to defining shape by local texture elements in naturalistic displays.) Precisely what the boundary conditions are on these findings remains to be determined. Because our dot stimuli are small, sparse, and hence of low total contrast power, they may be particularly poor stimuli for a second-order motion system. Prazdny (1986) reported an example of 3D shape from second-order motion stimuli (which do not effectively stimulate first-order mechanisms) for very simple (4 bend) wide wire figures. The wires were depicted by dense random dynamic noise against a background of dense static noise. His shapes were very simple, non-surface shapes, and were not edited to exclude 2D information about identity. However, his thick wires are a better stimulus (than our dots) for a second-order system due to the large spatial scale.

In a subsequent paper (Landy, Sperling, Doshier, & Perkins, 1988), we examine kinetic depth stimuli that are statistically invisible to Fourier detectors. We use various different stimulus tokens (dots, disks, wires) and backgrounds (gray, static random noise), as well as polarity alternation of standard stimuli. For large-scale tokens, polarity alternation is very

damaging, but some residual above-chance 3D shape identification appears to be possible. That investigation also supports and generalizes the conclusion that the *primary* substrate of shape identification is strong first-order motion information for stimuli which require analysis of motion in a number of regions simultaneously. However, appropriately constructed displays, which provide a high power stimulus to the second-order motion systems, may support reduced, but above-chance 3D shape analysis.

ACKNOWLEDGEMENT

This work was supported by Office of Naval Research, Grant N00014-85-K-007 and by AFOSR, Life Science Directorate, Visual Information Processing Program, Grants No. AFOSR 85-0364 and 88-0140.

FOOTNOTES

1. Preliminary reports of these experiments are contained in Landy, Sperling, Doshier, & Perkins (1987), Landy, Sperling, Perkins, & Doshier (1987), and Doshier, Landy, & Sperling (1988).

2. The number of dots actually varied slightly from 300 due to sampling of dots at or near the windowed edges.

3. The linearization of the monitor depended on the average intensification level. To equate light and dark dots required calibration on the same gray-level, and with display conditions as closely related to the actual displays as possible. A regular grid of one in nine pixels was nominally assigned the dark intensity and the remaining pixels assigned the gray background level. The decrement (in cd/m^2) relative to a uniform field of background intensity was equated to the increment when one in nine pixels were assigned the light intensity on a gray background level. One in nine pixels is an approximation to the sparse displays of the actual stimuli, while still providing stable measurements with an UDT-161CRT photometer. The increment in intensification due to each stimulus dot (in $\mu\text{cd/dot}$) was computed from the field increment. Although a stimulus dot is nominally one pixel, our calibrations show that intensification affects neighboring pixels via the point spread function of the monitor and

phosphor nonlinearities.

4. In the display of a transparent cylinder filled with dots, rotating around a central vertical axis and translating upward, dots viewed through the middle of the cylinder have a greater range of lateral motion velocities and dots at the 2D edges have a smaller range of velocities; In Williams and Phillips' random flow field, there is a wide range of velocities throughout the display. However, at the edges, dots disappear and re-appear; this scintillation (as in Experiment 2) reduces the magnitude of perceived depth; mean lateral velocity in both areas is zero. The effective flow fields for these differently constructed stimuli actually are quite similar.

5. At certain moments during the rotation, dots on bumps move opposite to ground dots, and at other moments dots on depressions move opposite to ground dots. To solve the task by motion direction only would require sampling at least 3 frames. That is, to observe any motion at all, requires two frames. Since there are only two categories of motion-direction response, from the motion observed in the first two frames, only two categories of dots could be observed (e.g., left or rightward moving). By observing a third frame, some of the dots that were categorized together in the first two frames could be differentiated (e.g., initially leftward, then rightward) and this could be used, in principle, to set up the three categories of dots (forward, center, behind) needed to solve the 3D shape discrimination task. However, we show (Landy, et al., 1988) that two frames suffice for accurate performance. This means that at least three (moving leftward, moving rightward, not moving) and probably more categories of velocity information are available. Therefore, for the present discussion, we can assume that our 3D shape identification task has access to three-category velocity information; this velocity information obtained simultaneously from (at least) six locations would suffice to solve the task.

6. To prove that the relevant cue for discriminating bump heights is speed, possible alternative cues, such as distance traversed and the configuration at the point of rotation reversal must be irrelevantly varied so that they can not become artifactual cues.

REFERENCES

- Adelson, E. H., & Bergen, J. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America, A2*, 284-299.
- Adelson, E. H., & Bergen, J. R. (1986). The extraction of spatio-temporal energy in human and machine vision. *Proceedings of Workshop on Motion: Representation and Analysis*, IEEE Computer Society # 696, 151-155.
- Anderson, G. J., & Braunstein, M. L. (1983). Dynamic occlusion in the perception of rotation in depth. *Perception and Psychophysics, 34*, 356-362.
- Anstis, S. M. (1970). Phi movement as a subtraction process. *Vision Research, 10*, 1411-1430.
- Anstis, S. M., & Rogers, B. J. (1975). Illusory reversal of visual depth and movement during changes of contrast. *Vision Research, 15*, 957-961.
- Ball, K., & Sekuler, R. (1979). Masking of motion by broad band and filtered directional noise. *Perception and Psychophysics, 26*, 206-214.
- Braddick, O. (1973). The masking of apparent motion in random-dot patterns. *Vision Research, 13*, 355-369.
- Braddick, O. (1974). A short range process in apparent motion. *Vision Research, 14*, 519-527.
- Braunstein, M. L. (1962). Depth perception in rotating dot patterns: Effects of numerosity and perspective. *Journal of Experimental Psychology, 64*, 415-420.

- Braunstein, M. L., Hoffman, D. D., Shapiro, L. R., Anderson, G. J., & Bennett, B. M. (1987). Minimum points and views for the recovery of three-dimensional structure. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 335-343.
- Burr, D. C., & Ross, J. (1982). Contrast sensitivity at high velocities. *Vision Research*, 22, 479-484.
- Burt, P., & Sperling, G. (1981). Time, distance, and feature trade-offs in visual apparent motion. *Psychological Review*, 88, 171-195.
- Chubb, C., & Sperling, G. (1987). Drift-balanced random stimuli: A general basis for studying nonFourier motion perception. *Investigative Ophthalmology and Visual Science (Supplement)*, 28, 233.
- Chubb, C., & Sperling, G. (1988a). Processing stages in non-Fourier motion perception. *Investigative Ophthalmology and Visual Science (Supplement)*, 29, 266.
- Chubb, C., & Sperling, G. (1988b). Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception. *Journal of the Optical Society of America A: Optics and Image Science*, 5, 1986-2006.
- Chubb, C., & Sperling, G. (1989a). Second-order motion perception: Space-time separable mechanisms. *Proceedings: 1989 IEEE Workshop on Motion*. Washington, D.C: IEEE Computer Society Press, in press.
- Chubb, C., & Sperling, G. (1989b). Two motion perception mechanisms revealed by distance driven reversal of apparent motion. *Proceedings of the National Academy of Sciences, USA*, 86, in press.

- Cocksins, W. F. (1980). Perception of surface slant and edge labels from optical flow: a computational approach. *Perception*, 9, 253-269.
- van Doorn, A. J., & Koenderink, J. J. (1982). Spatial properties of the visual detectability of moving spatial white noise. *Experimental Brain Research*, 45, 189-195.
- Dosher, B. A., Landy, M. S., & Sperling, G. (1989). Ratings of kinetic depth in multi-dot displays. *Journal of Experimental Psychology: Human Perception and Performance*, in press.
- Dosher, B. A., Landy, M. S., & Sperling, G. (1988). The kinetic depth effect and optic flow. I. 3D Shape from Fourier motion. *Mathematical Studies in Perception and Cognition*, 88-4, NYU Report Series.
- Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving images. *Computer Graphics and Image Processing*, 9, 301-315.
- Foster, D. H. (1969). The response of the human visual system to moving spatially-periodic patterns. *Vision Research*, 9, 577-590.
- Foster, D. H. (1971). The response of the human visual system to moving spatially-periodic patterns: Further analysis. *Vision Research*, 11, 57-81.
- Green, B. F. (1961). Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology*, 62, 272-282.
- Green, M. (1983). Contrast detection and direction discrimination of drifting gratings. *Vision Research*, 23, 281-289.

- Harris, M. G. (1986). The perception of moving stimuli: A model of spatiotemporal coding in human vision. *Vision Research*, 26, 1281-1287.
- Heeger, D. J. (1987). A model for the extraction of image flow. *Journal of the Optical Society of America A*, 4, 1455-1471.
- Hoffman, D. D. (1982). Inferring local surface orientation from motion fields. *Journal of the Optical Society of America*, 72, 888-892.
- Hoffman, D. D., & Bennett, B. M. (1985). Inferring the relative three-dimensional positions of two moving points. *Journal of the Optical Society of America A*, 2, 350-353.
- Horn, B. K. P., & Schunk, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17, 185-203.
- Koenderink, J. J., & van Doorn, A. J. (1986). Depth and shape from differential perspective in the presence of bending deformations. *Journal of the Optical Society of America A*, 3, 242-249.
- Krauskopf, J. (1980). Discrimination and detection of changes in luminance. *Vision Research*, 20, 671-677.
- Landy, M. S., Doshier, B. A., Sperling, G., & Perkins, M. E. (1988). The kinetic depth effect and optic flow. II. Fourier and non-Fourier motion. *Mathematical Studies in Perception and Cognition*, 88-4, NYU Report Series.
- Landy, M. S., Sperling, G., Doshier, B. A., & Perkins, M. E. (1987). From what kind of motions can structure be inferred? *Investigative Ophthalmology and Visual Science (Supplement)*, 28, 233.

- Landy, M. S., Sperling, G., Perkins, M. E., & Doshier, B. A. (1987). Perception of complex shape from optic flow. *Journal of the Optical Society of America A: Optics and Image Science*, 1987, 4, No. 13, P95.
- Limb, J. O., & Murphy, J. A. (1978). Estimating the velocity of moving images in television signals. *Computer Graphics and Image Processing*, 4, 311-327.
- Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of Image Understanding Workshop*, 1221-130.
- Marr, D., & Ullman, S. (1981). Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London, B*, 211, 151-180.
- Nakayama, K. (1985). Biological image motion processing: A review. *Vision Research*, 25, 625-660.
- Patel, A. S., & Jones, R. W. (1968). Increment and decrement visual thresholds. *Journal of the Optical Society of America*, 58, 696-699.
- Prazdny, K. (1987). Three-dimensional structure from long-range apparent motion. *Perception*, 15, 619-625.
- Rashbass, C. (1970). The visibility of transient changes of luminance. *Journal of Physiology*, 210, 165-186.
- Reichardt, W. (1957). Autokorrelationsauswertung als funktionsprinzip des zentralnervensystems. *Z. Naturforschung*, 12b, 447-457.
- Rogers, B. J., & Anstis, S. M. (1975). Reversed depth from positive and negative

stereograms. *Perception*, 4, 193-201.

Roufs, J. A. J. (1974). Dynamic properties of vision - VI. Stochastic threshold fluctuations and their effect on flash-to-flicker sensitivity ratio. *Vision Research*, 14, 871-888.

van Santen, J. P. H., & Sperling, G. (1984a). A temporal covariance model of motion perception. *Journal of the Optical Society of America A*, 1, 451-473.

van Santen, J. P. H., & Sperling, G. (1984b). Applications of a Reichardt-type model to two-frame motion. *Investigative Ophthalmology and Visual Science (Supplement)*, 25, 14.

van Santen, J. P. H., & Sperling, G. (1985). Elaborated Reichardt detectors. *Journal of the Optical Society of America A*, 2, 300-321.

Short, A. D. (1966). Decremental and incremental thresholds. *Journal of Physiology*, 185, 646-654.

Sperling, G. (1976). Movement perception in computer-driven visual displays. *Behavior, Research, Methods and Instrumentation*, 8, 144-151.

Sperling, G., Landy, M. S., Doshier, B. A., & Perkins, M. E. (1989). The kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception and Performance*, in press.

Ullman, S. (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.

Ullman, S. (1985). Maximizing rigidity: the incremental recovery of 3-D structure from rigid and non-rigid motion. *Perception*, 13, 255-274.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental*

Psychology, 45, 205-217.

Watson, A. B. (1986). Temporal sensitivity. In *Handbook of Perception and Human Performance, Volume I: Sensory Processes and Perception* (K. R. Boff, L. Kaufman, & J. P. Thomas, Eds.). New York: John Wiley and Sons.

Watson, A. B., & Ahumada, A. J., Jr. (1983). A look at motion in the frequency domain. *NASA Technical Memorandum* 84352.

Watson, A. B., & Ahumada, A. J., Jr. (1984). A model of how humans sense image motion. *Investigative Ophthalmology and Visual Science (Supplement)*, 25, 14.

Watson, A. B., & Ahumada, A. J., Jr. (1985). Model of human visual-motion sensing. *Journal of the Optical Society of America A*, 1, 322-342.

Watson, A. B., Ahumada, A. J., Jr., & Farrell, J. E. (1986). Window of visibility: A psychophysical theory of fidelity in time-sampled visual motion displays. *Journal of the Optical Society of America A*, 3, 300-307.

Watson, A. B., Thompson, P. G., Murphy, B. J., & Nachmias, J. (1980). Summation and discrimination of gratings moving in opposite directions. *Vision Research*, 20, 341-347.

Williams, D., & Phillips, G. (1986). Structure from motion in a stochastic display. *Journal of the Optical Society of America A*, 3, 30-31.

Williams, D., & Phillips, G. (1987). Rigid 3-D percept from stochastic 1-D motion. *Journal of the Optical Society of America A*, 4, 48.

FIGURE LEGENDS

Figure 1. A schematic illustration of an Elaborated Reichardt Detector (van Santen & Sperling, 1985), one implementation of a spatio-temporal motion analyzer. Image intensity at location A at time t is correlated (multiplied) by image intensity at location B at time $t+\Delta t$ (left half-detector). Similarly, image intensity at location B at time t is correlated (multiplied) by image intensity at location A at time $t+\Delta t$ (right half-detector). These correlation values are temporally integrated over some time domain T , and compared (subtracted) to yield a direction-of-motion signal for that detector. Orientation and velocity tuning are determined by the selection of receptive fields I_A and I_B and Δt . Spatial scale is determined by the spatial function which senses image intensity. Outputs of populations of such detectors of various scales, locations, and velocity tuning must be integrated with subsequent decision rules. Further elaborations are required to construct velocity sensors.

Figure 2. (a) Schematic illustration of a simple spatio-temporal sensor operating on a moving white dot on a gray background. One dimension of space x , and time t are represented. The center (solid ellipse) has a weight of $+1$; each of the flanks (dotted ellipse) has a weight of $-1/2$. The geometry and orientation of the hypothetical receptive field represent the preference for a particular spatial scale, direction, and velocity. (b) Same sensor as (a) operating on a stimulus with interleaved gray frames, and a second sensor sensitive to the opposite velocity. The magnitude of the stimulation of the center of sensor 1 equals the combined magnitude of the stimulation of the two flanks of sensor 2. At this scale, there is equal evidence for both orientations, i.e., both velocities. (c) Same sensor as (a) operating on a stimulus with tokens alternating polarity above and below the gray background level. Sensor 1 receives oppositely signed inputs in its center and has a weak output. Sensor 2 receives inputs in its surround opposite in sign from those in its center and therefore has a large output. Alternating polarity

yields strong evidence for orientation from upper right to lower left, i.e., for motion opposite to the direction in (a).

Figure 3. (a) Illustration of the upward and downward pointing triangular layout of peak and valley locations in the shape lexicon. Members of the lexicon may have either the upward or downward layout, and either a peak, valley, or ground value at each of the three locations. (b) Examples of a number of shapes in the shape lexicon as defined by a rectangular grid spline over peaks and valleys. Actual stimuli consisted of parallel projections of dots sprinkled over these shapes undergoing sinusoidal rotation. Subjects were required to identify the shape and the direction of rotation. (c) Schematic illustration of the shape identification displays with rotation. (d) A single frame of a 2D image sequence for the shape identification task.

Figure 4. Shape identification performance for normal displays with and without density cues, and for the density only displays. Performance range is from 0 to 100%, with a guessing baserate of 1.9%. The three panels show data for individual subjects.

Figure 5. Shape identification performance for normal displays, alternating gray frame displays, alternating polarity displays and a number of control displays. Performance range is from 0 to 100%, with a guessing baserate of 1.9%. The three panels show data for individual subjects. (Contrast equated condition unavailable for subject CFS.)

Figure 6. (a) Illustration of the construction of two-frame lifetime displays, as well as standard construction. In the top panel, sampled dots remain visible in all frames of the display. In the standard 30-frame condition, control for density cues actually introduced 5% scintillation, for an expected lifetime of 20 frames. In the bottom panel, sampled dots remain only for two frames, and are then replaced by another sample. (b) Percent shape identification for three subjects in each of the lifetime display conditions. Guessing baserate is 1.9%. Shape

identification is little affected by the lifetime manipulation. The small decline may be a consequence of scintillation not loss of trajectory information.

Figure 7. (a) Illustration of the two-interval forced choice (2IFC) paradigm for the motion visibility task. Subjects judged which 1 sec interval contained a stimulus, and which interval was blank. (b) Percent detection of a planar motion display in the 2IFC performance. Detection is measured for normal and polarity alternation image sequences as a function of dot intensity (expressed as a percentage of a standard intensity). Guessing baserate is 50%.

Figure 8. (a) Schematic illustration of the motion direction discrimination task. Outer dots were dynamic noise, dots in the central patch drifted left or right at 0.35 deg/sec. Subjects judged the direction of motion of dots in the central patch. (b) Percent correct discrimination of the direction of motion in the 2D motion-direction display. Discrimination is shown as a function of the intensity increment (as a percent of the "standard" intensity increment), of the stimulus dots on a gray background. The intensity increment where the dashed line-and-arrow intersects the performance line for standard displays equates standard (at reduced intensities) and polarity alternation displays (at standard intensities). The guessing baserate is 50%. Panels show the data of different subjects.

Figure 9. (a) Schematic illustration of the nine-location forced-choice (9LFC) motion segmentation display. Subjects judged the location of the single patch moving opposite in direction to the other eight. (b) Percent correct location judgment for the 9LFC task for standard and alternating polarity displays for two subjects. Guessing baseline is 11.1% (1 in 9).

Figure 10. Stimulus representations and corresponding Fourier energy spectra typical of various display conditions. (a,c,e,g,i,k) Each stimulus representation depicts 1.07 sec of planar

motion of a single dot moving at a rate of 0.35 deg/sec. The abscissa is (horizontal) spatial location, and the ordinate is time. The representation assumes spatial resolution of 60 cycles per degree of visual angle, and temporal resolution of 60 Hz. The stimulus is either light or dark increments or decrements on a gray background. (b,d,f,h,j,l) The corresponding Fourier spectra are shown on ω_x (abscissa), ω_t (ordinate) axes. The inner boxes represent the window of visibility, assumed to resolve less than or equal to 30 cycles per degree and less than or equal to 30 Hz. The upper left (or lower right) quadrant of the spectra represent power at (ω_x, ω_t) consistent with the intended direction of motion. The upper right (or lower left) quadrant of the spectra represent power at (ω_x, ω_t) consistent with the unintended direction of motion. The representation and spectrum for the "standard" stimulus are shown in (a,b), for the half-contrast standard stimulus in (c,d), for the alternating-gray stimulus in (e,f), for the alternating-polarity stimulus in (g,h), for the alternating contrast 2:1 in (i,j), and for the alternating contrast 1.5:0.5 in (k,l).

Figure 11. A schematic illustration of the kinds of information required in order to perform each of the experimental tasks. The simple 2IFC detection task may reflect the output of non-motion systems in a single location. The 2AFC discrimination of motion direction task requires the output of a motion direction mechanism in a single location. The 9LFC motion segmentation task requires the output of motion direction mechanisms in a number of locations nearly simultaneously. The 3D shape task requires direction and speed information from a number of locations nearly simultaneously.

Figure 12. The relation between 3D shape identification performance and computed *net directional power* DP within the window of visibility and above a threshold ϵ . Solid circles on the abscissa are values of DP computed from the spectra in Figure 10, panels (b), (d), etc., for an ϵ of .12 \times the maximum power value in the spectrum of the standard stimulus. Open

circles on the abscissa are the values of DP computed for an ϵ of 0. (The rank order of conditions under the two computations is the same.) The 3D shape identification performance is monotone with DP for all reasonable values of $\epsilon \geq 0$.

Table 1.
Display Types for Experiments 1, 2, and 3

Task: Large Lexicon Shape Identification					
Display	Motion Cue(a)	Density Cue(b)	Rotation Speed(c)	Intensity +/- Increments(d)	Dot Lifetime(e)
Experiment 1 (Main)					
1. With Density	3D	Y	Standard	1 : 1	30
2. Standard	3D	N	Standard	1 : 1	≤30
3. With Density	3D	Y	Half	1 : 1	30
4. Standard	3D	N	Half	1 : 1	≤30
5. Alternating Polarity	3D	N	Standard	1 : -1	≤30
6. Alternating Polarity	3D	N	Standard	.5 : -.5	≤30
7. Alternating Gray	3D	N	Half	1 : 0	≤30
8. Alternating Gray	3D	N	Standard	1 : 0	≤30
9. Alternating Contrast	3D	N	Standard	2 : 1	≤30
10. Alternating Contrast	3D	N	Standard	1.5 : .5	≤30
11. Density Only	Random	Y	Standard	1 : 1	1
Experiment 2 (Equated Contrast)					
12. Standard	3D	N	Standard	V : V	≤30
Experiment 3 (Lifetimes)					
2. Standard	3D	N	Standard	1 : 1	≤30
13. 3 Frame	3D	N	Standard	1 : 1	3
14. 2 Frame	3D	N	Standard	1 : 1	2

Notes to Table 1

- (a) 3D motion cues refers to 2D projections of 3D moving stimuli. Random refers to random motion correspondences arising from uncorrelated new dot samples on each frame.
- (b) Dot-density cues removed by minimal (<5%) dot scintillation.
- (c) Standard rotation speed: was ± 25 deg sinusoidal rotation per 30 new frames; 15 new frames per sec with 4 sync cycles per new frame; Half rotation speed: ± 25 deg sinusoidal rotation per 30 new frames; 7.5 new frames per sec with 8 sync cycles per new frame (conditions 3, 4) or 15 new frames per sec with 4 sync cycles per new frame (condition 7) (see text).
- (d) The numbers code the increments or decrements in intensification of dots on a neutral gray background. 1 refers to $1 \times$ the standard increment level, and -1 refers to $1 \times$ the standard decrement level. The value to the left of the colon refers to dot intensification on odd frames; the value to the right to even frames. For example, 1:1 means dots received the same standard increments on all frames; 1:0 means dots received standard intensification on odd frames, and no intensification on even frames; etc. Gray background was between 31 and 38 cd/m^2 . Standard increments (and decrements) were between 13 and 21 extra (or fewer) μcd per dot. See the text for exact values for each subject. The value V refers to a fraction (<1) of standard increment intensity which equates non-alternating stimuli to alternating polarity stimuli for percent correct planar motion direction judgements (see Experiment 5). Intensities for V were between approximately .5-.6, or between 8 and 10 μcd per dot.
- (e) Lifetime refers to the number of new frames that the same dots on the 3D surface appear in during the stimulus sequence. Since the display sequences were 30 new frames long, a lifetime of 30 frames is maximal. The value ≤ 30 refers to nominal lifetime of 30 frames, subject to scintillation for density control. Conditions (13) and (14) resample one third and one half of the dots in the stimulus per frame, respectively, yielding scintillation values of 33% and 50%.

Table 2.
Display Types for Experiments 4, 5, and 6

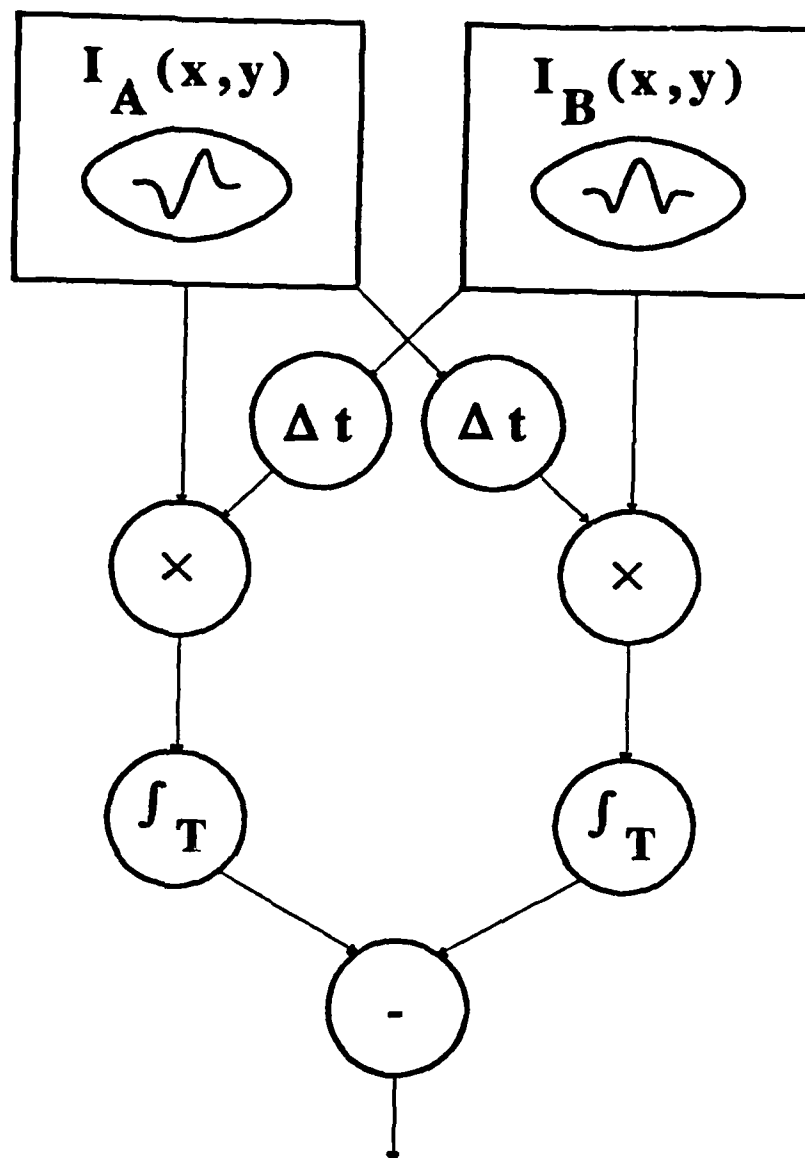
Planar Motion Experiments					
Display	Motion Cue(a)	Number of Patches(b)	Motion Direction	Intensity +/- Increments(c)	Task
Experiment 4 (Visibility)					
15-19. Standard	2D	1	L or R	5 Levels	Detection (2IFC)
20-24. Alternating Polarity	2D	1	L or R	+/-5 Levels	Detection (2IFC)
Experiment 5 (Motion Direction)					
25-29. Standard	2D	1	L or R	5 Levels	Direction (2AFC)
30-34. Alternating Polarity	2D	1	L or R	+/-5 Levels	Direction (2AFC)
Experiment 6 (Motion Segmentation)					
35. Standard	2D	9	8L/1R or 1L/8R	1 : 1	Odd Motion (9AFC)
36. Alternating Polarity	2D	9	8L/1R or 1L/8R	1 : -1	Odd Motion (9AFC)

Notes to Table 2

(a) 2D motion cue refers to uniform field motion of a random dot field in a larger background of neutral gray or of dynamic random dot noise. Planar motion was 1 pixel per new frame, 15 new frames per sec, or 4 sync cycles per new frame. See text for details.

(b) Patches were 48 x 48 pixels. Single patches were embedded in a larger background. The 9-patch displays were arranged in a 3 X 3 square grid.

(c) Dots were displayed as increments or decrements on a gray background. The intensities were varied as percentages of the standard increments and decrements, which are labeled as in Table 1. Variable intensity increments differed across subjects (see text).



Further Processing and Decision Rules

Figure 1. A schematic illustration of an Elaborated Reichardt Detector (van Santen & Sperling, 1985), one implementation of a spatio-temporal motion analyzer. Image intensity at location A at time t is correlated (multiplied) by image intensity at location B at time $t+\Delta t$ (left half-detector). Similarly, image intensity at location B at time t is correlated (multiplied) by image intensity at location A at time $t+\Delta t$ (right half-detector). These correlation values are temporally integrated over some time domain T , and compared (subtracted) to yield a direction-of-motion signal for that detector. Orientation and velocity tuning are determined by the selection of receptive fields I_A and I_B and Δt . Spatial scale is determined by the spatial function which senses image intensity. Outputs of populations of such detectors of various scales, locations, and velocity tuning must be integrated with subsequent decision rules. Further elaborations are required to construct velocity sensors.

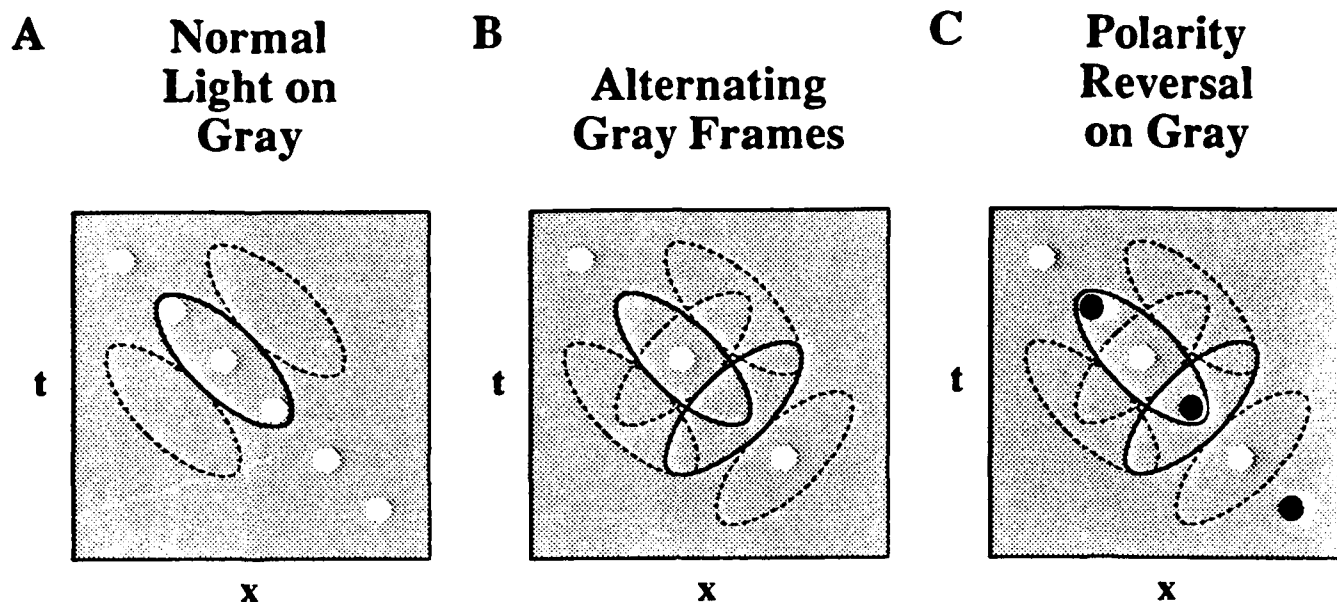


Figure 2. (a) Schematic illustration of a simple spatio-temporal sensor operating on a moving white dot on a gray background. One dimension of space x , and time t are represented. The center (solid ellipse) has a weight of $+1$; each of the flanks (dotted ellipse) has a weight of $-1/2$. The geometry and orientation of the hypothetical receptive field represent the preference for a particular spatial scale, direction, and velocity. (b) Same sensor as (a) operating on a stimulus with interleaved gray frames, and a second sensor sensitive to the opposite velocity. The magnitude of the stimulation of the center of sensor 1 equals the combined magnitude of the stimulation of the two flanks of sensor 2. At this scale, there is equal evidence for both orientations, i.e., both velocities. (c) Same sensor as (a) operating on a stimulus with tokens alternating polarity above and below the gray background level. Sensor 1 receives oppositely signed inputs in its center and has a weak output. Sensor 2 receives inputs in its surround opposite in sign from those in its center and therefore has a large output. Alternating polarity yields strong evidence for orientation from upper right to lower left, i.e., for motion opposite to the direction in (a).

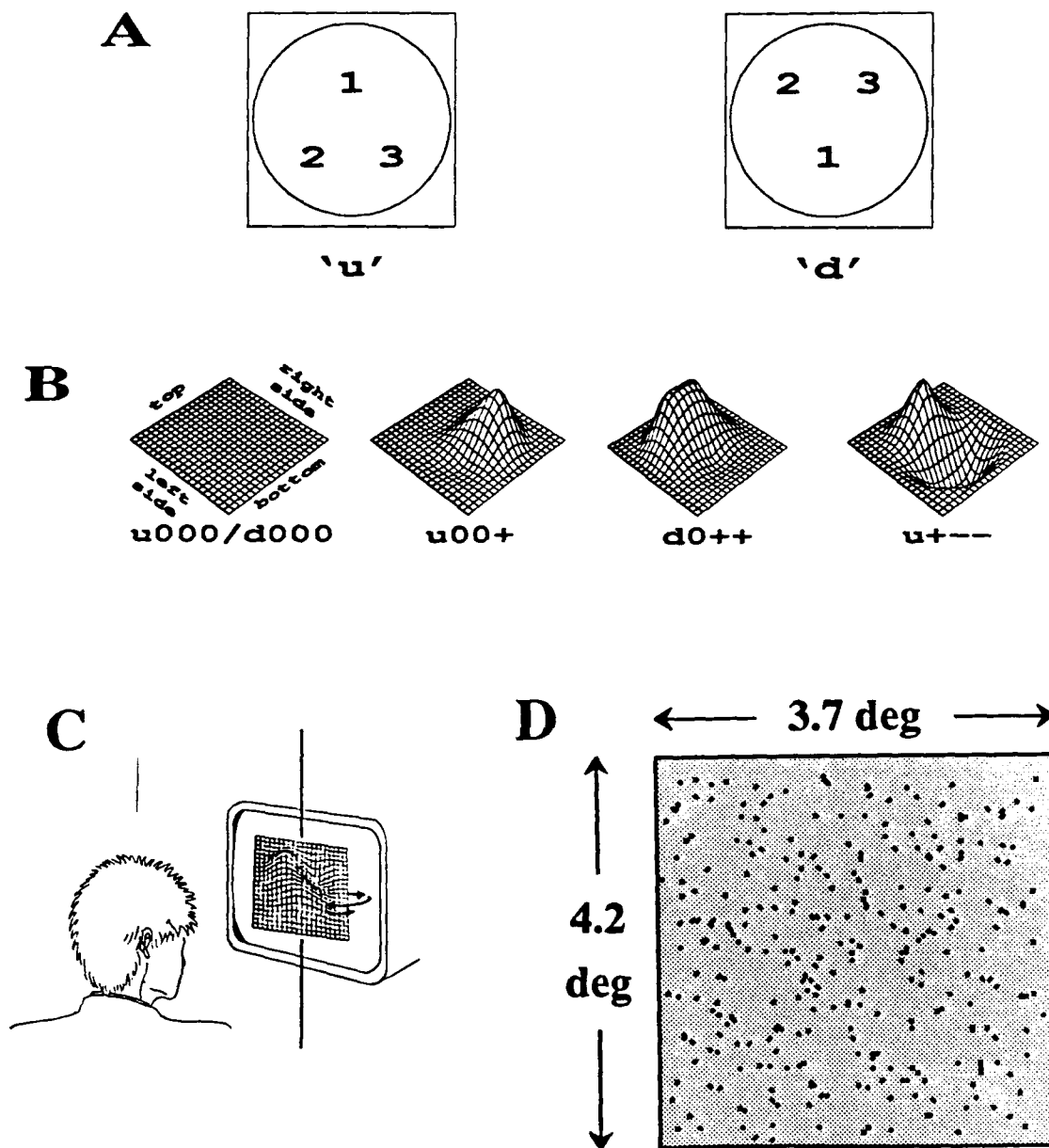


Figure 3. (a) Illustration of the upward and downward pointing triangular layout of peak and valley locations in the shape lexicon. Members of the lexicon may have either the upward or downward layout, and either a peak, valley, or ground value at each of the three locations. (b) Examples of a number of shapes in the shape lexicon as defined by a rectangular grid spline over peaks and valleys. Actual stimuli consisted of parallel projections of dots sprinkled over these shapes undergoing sinusoidal rotation. Subjects were required to identify the shape and the direction of rotation. (c) Schematic illustration of the shape identification displays with rotation. (d) A single frame of a 2D image sequence for the shape identification task.

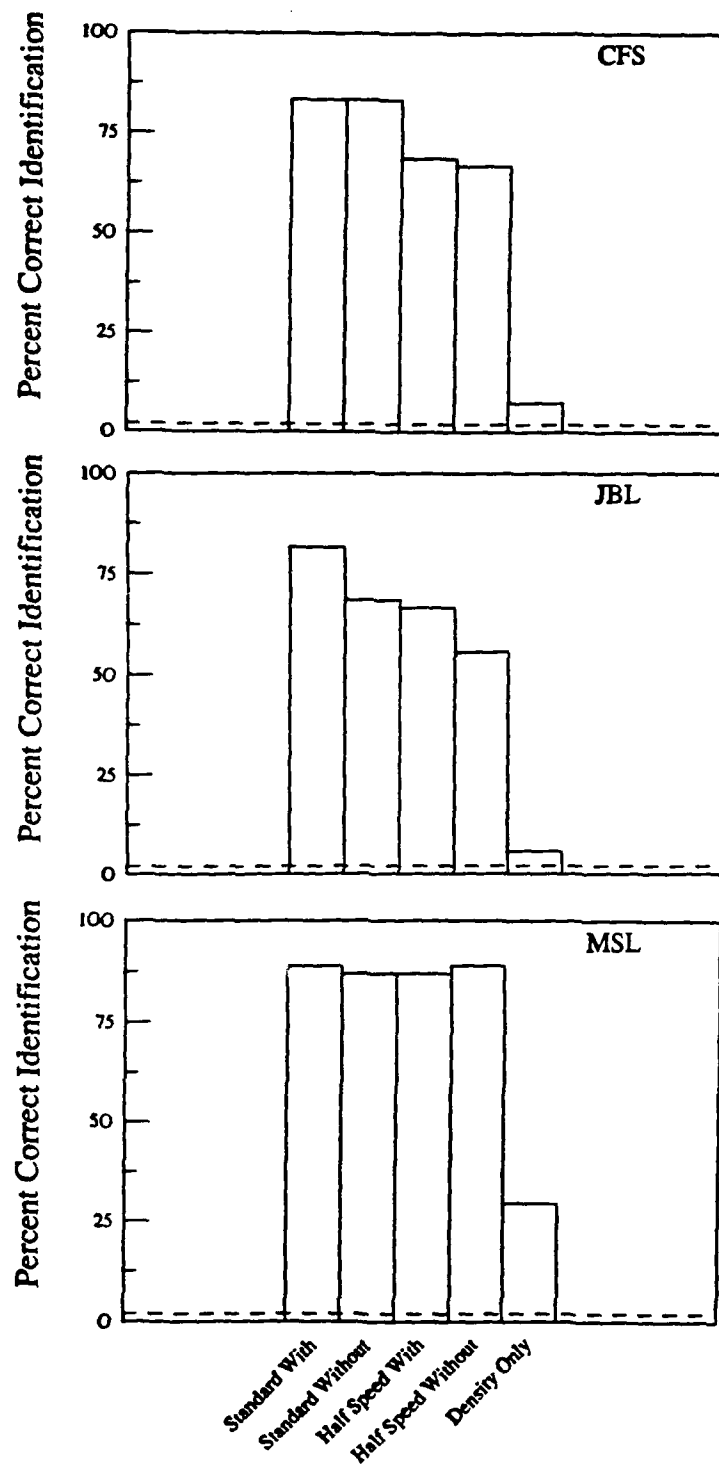


Figure 4. Shape identification performance for normal displays with and without density cues, and for the density only displays. Performance range is from 0 to 100%, with a guessing baserate of 1.9%. The three panels show data for individual subjects.

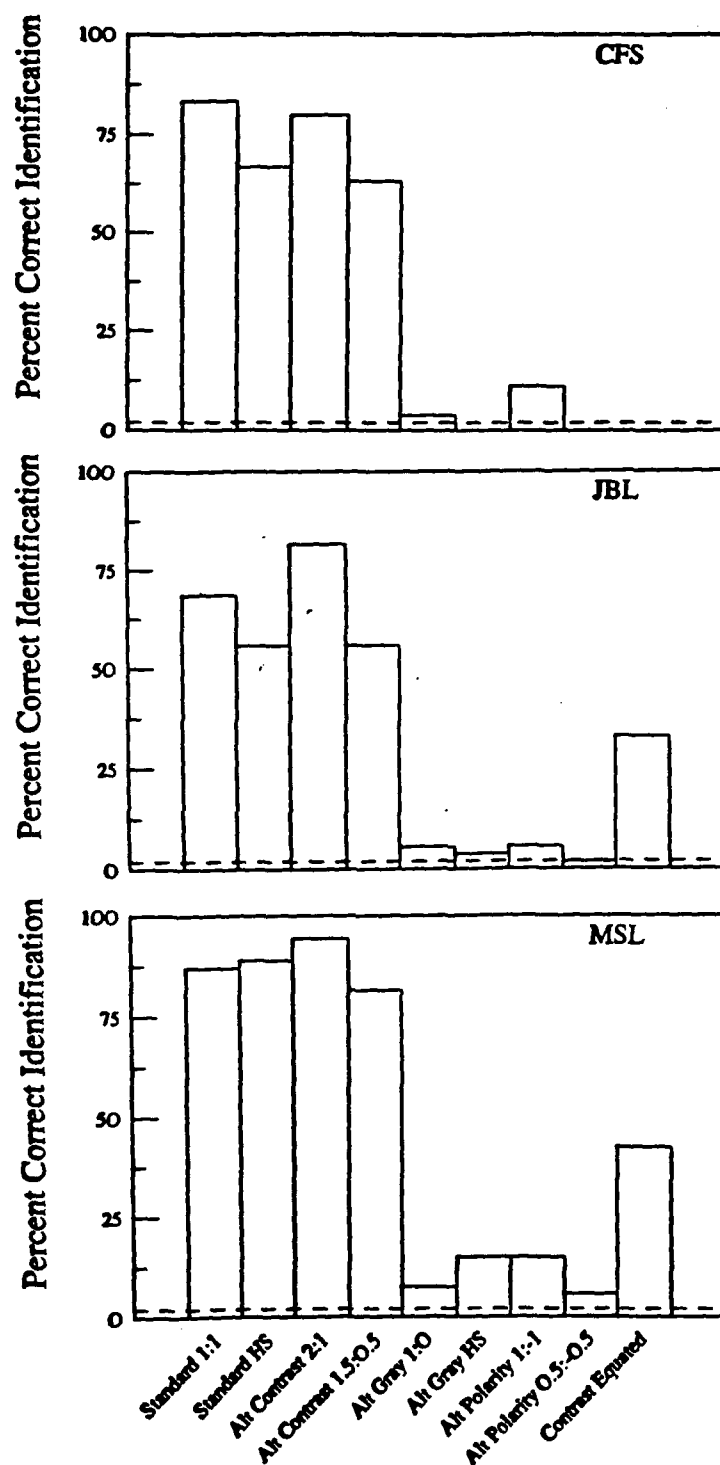
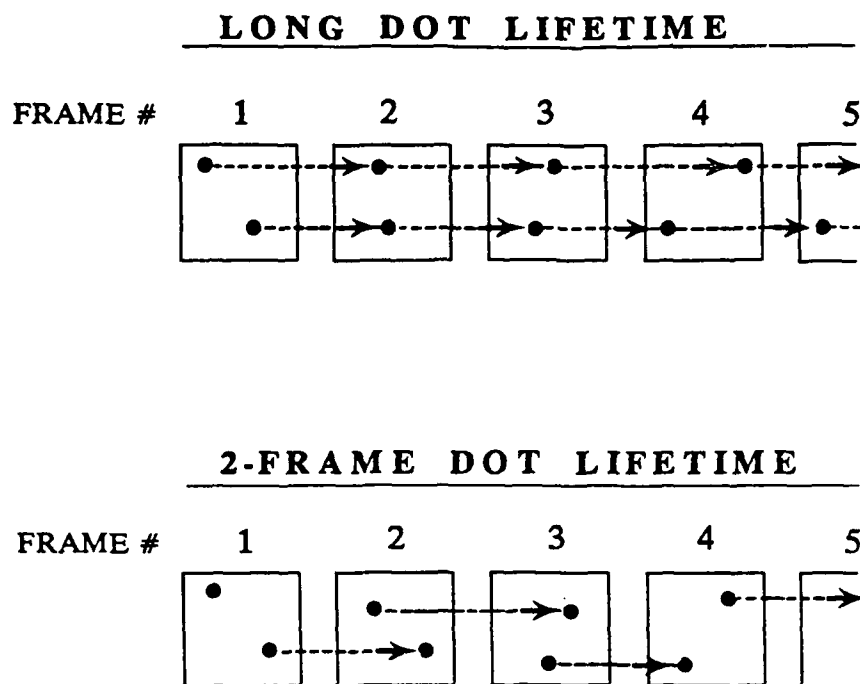


Figure 5. Shape identification performance for normal displays, alternating gray frame displays, alternating polarity displays and a number of control displays. Performance range is from 0 to 100%, with a guessing base rate of 1.9%. The three panels show data for individual subjects.

(a)



(b)

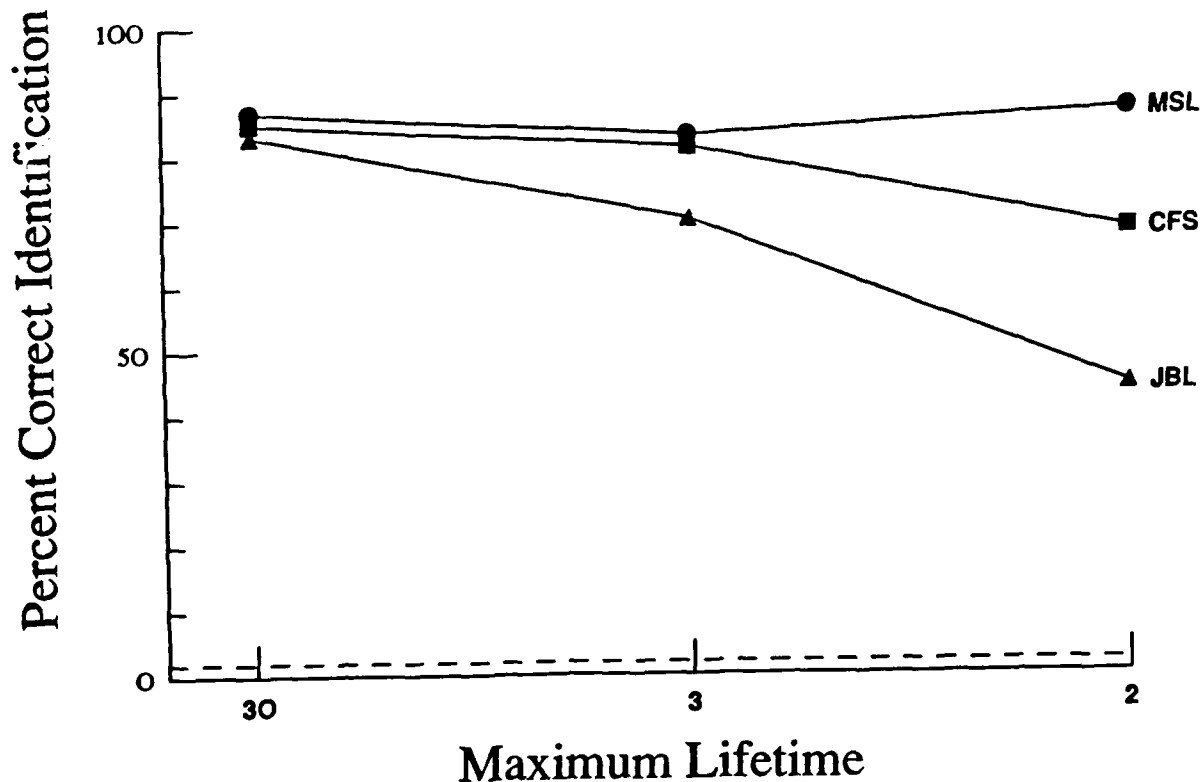


Figure 6. (a) Illustration of the construction of two-frame lifetime displays, as well as standard construction. In the top panel, sampled dots remain visible in all frames of the display. In the standard 30-frame condition, control for density cues actually introduced 5% scintillation, for an expected lifetime of 20 frames. In the bottom panel, sampled dots remain only for two frames, and are then replaced by another sample. (b) Percent shape identification for three subjects in each of the lifetime display conditions. Guessing base rate is 1.9%. Shape identification is little affected by the lifetime manipulation. The small decline may be a consequence of scintillation not loss of trajectory information.

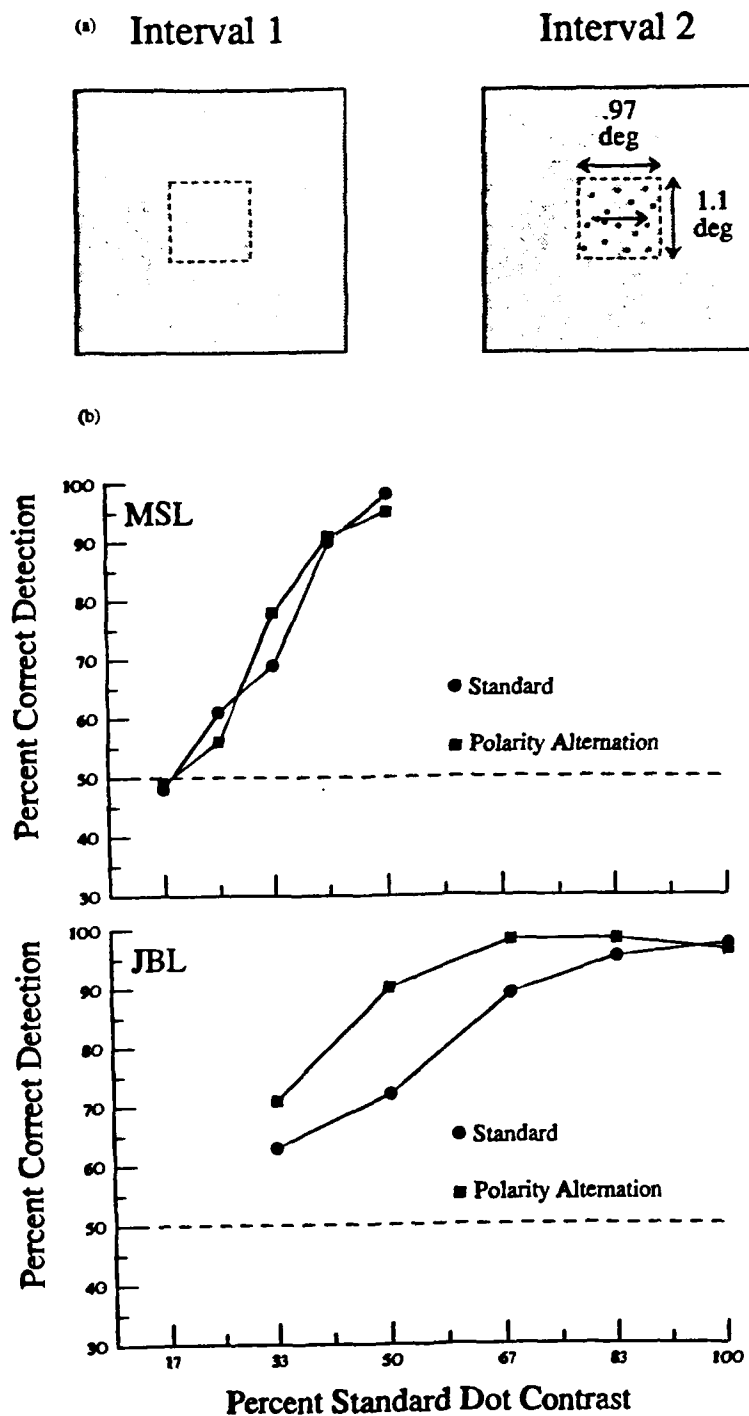


Figure 7. (a) Illustration of the two-interval forced choice (2IFC) paradigm for the motion visibility task. Subjects judged which 1 sec interval contained a stimulus, and which interval was blank. (b) Percent detection of a planar motion display in the 2IFC performance. Detection is measured for normal and polarity alternation image sequences as a function of dot intensity (expressed as a percentage of a standard intensity). Guessing base rate is 50%.

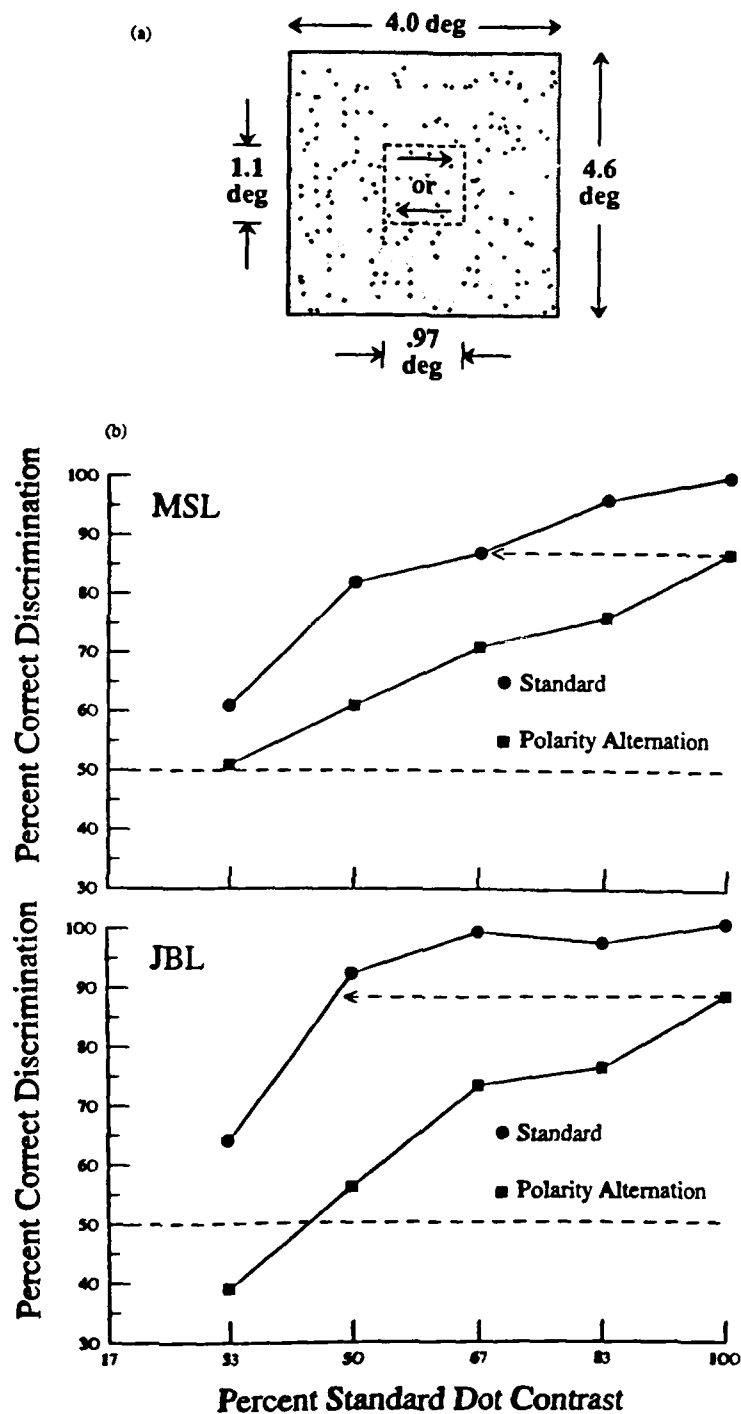


Figure 8. (a) Schematic illustration of the motion direction discrimination task. Outer dots were dynamic noise, dots in the central patch drifted left or right at 0.35 deg/sec. Subjects judged the direction of motion of dots in the central patch. (b) Percent correct discrimination of the direction of motion in the 2D motion-direction display. Discrimination is shown as a function of the intensity increment (as a percent of the "standard" intensity increment), of the stimulus dots on a gray background. The intensity increment where the dashed line-and-arrow intersects the performance line for standard displays equates standard (at reduced intensities) and polarity alternation displays (at standard intensities). The guessing baserate is 50%. Panels show the data of different subjects.

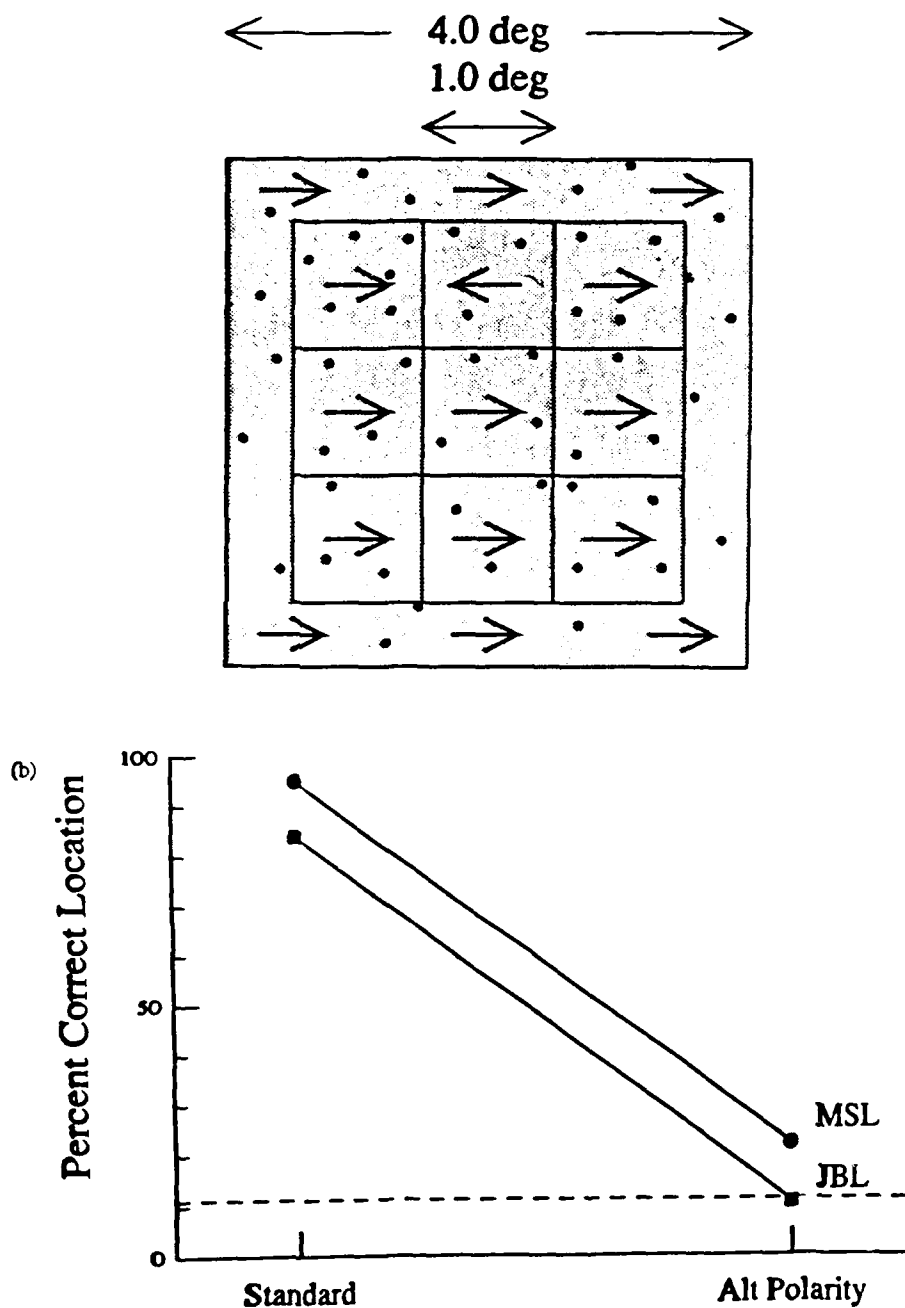


Figure 9. (a) Schematic illustration of the nine-location forced-choice (9LFC) motion segmentation display. Subjects judged the location of the single patch moving opposite in direction to the other eight. (b) Percent correct location judgment for the 9LFC task for standard and alternating polarity displays for two subjects. Guessing baseline is 11.1% (1 in 9).

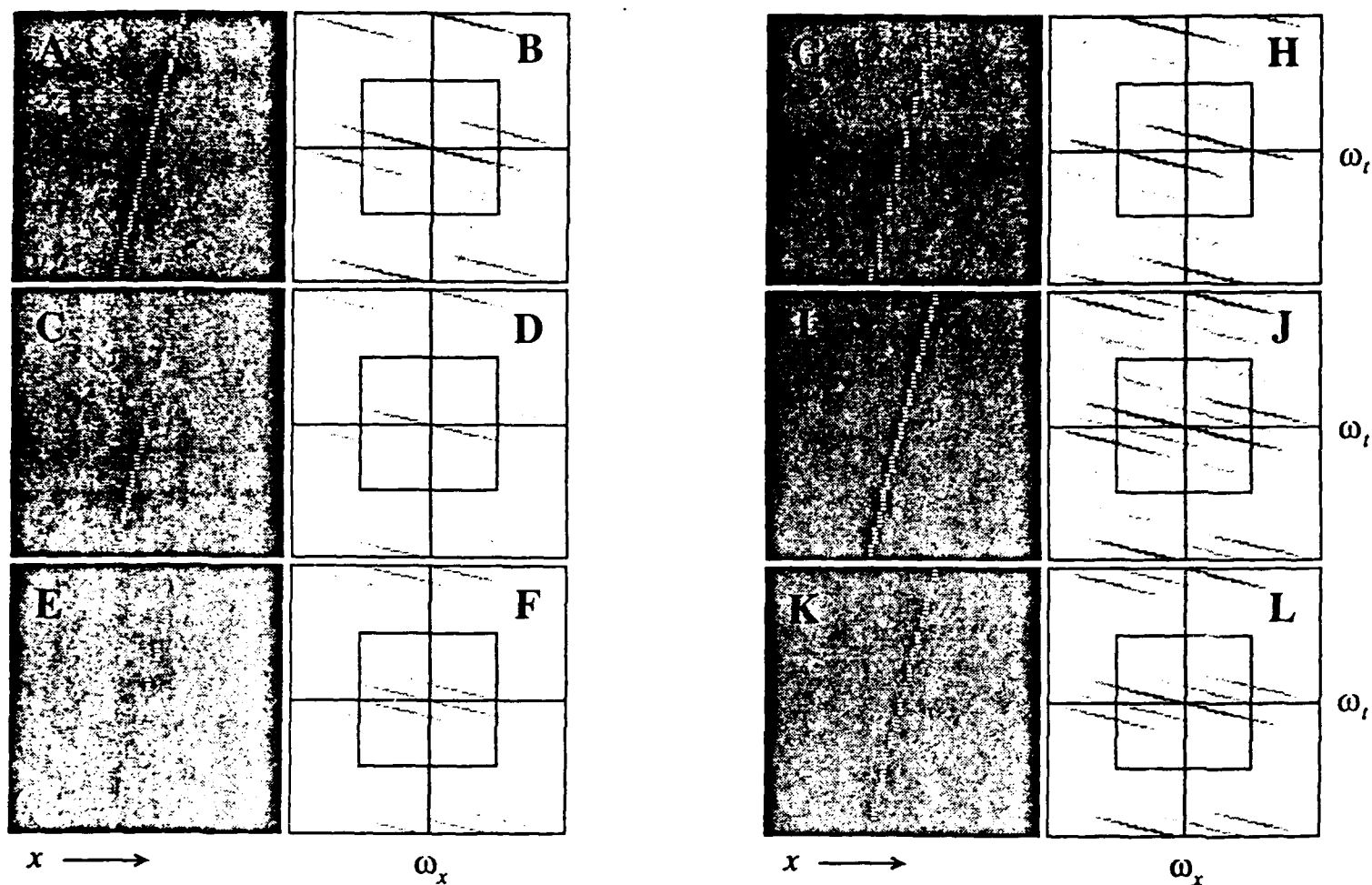


Figure 10. Stimulus representations and corresponding Fourier energy spectra typical of various display conditions. (a,c,e,g,i,k) Each stimulus representation depicts 1.07 sec of planar motion of a single dot moving at a rate of 0.35 deg/sec. The abscissa is (horizontal) spatial location, and the ordinate is time. The representation assumes spatial resolution of 60 cycles per degree of visual angle, and temporal resolution of 60 Hz. The stimulus is either light or dark increments or decrements on a gray background. (b,d,f,h,j,l) The corresponding Fourier spectra are shown on ω_x (abscissa), ω_t (ordinate) axes. The inner boxes represent the window of visibility, assumed to resolve less than or equal to 30 cycles per degree and less than or equal to 30 Hz. The upper left (or lower right) quadrant of the spectra represent power at (ω_x, ω_t) consistent with the intended direction of motion. The upper right (or lower left) quadrant of the spectra represent power at (ω_x, ω_t) consistent with the unintended direction of motion. The representation and spectrum for the "standard" stimulus are shown in (a,b), for the half-contrast standard stimulus in (c,d), for the alternating-gray stimulus in (e,f), for the alternating-polarity stimulus in (g,h), for the alternating contrast 2:1 in (i,j), and for the alternating contrast 1.5:0.5 in (k,l).

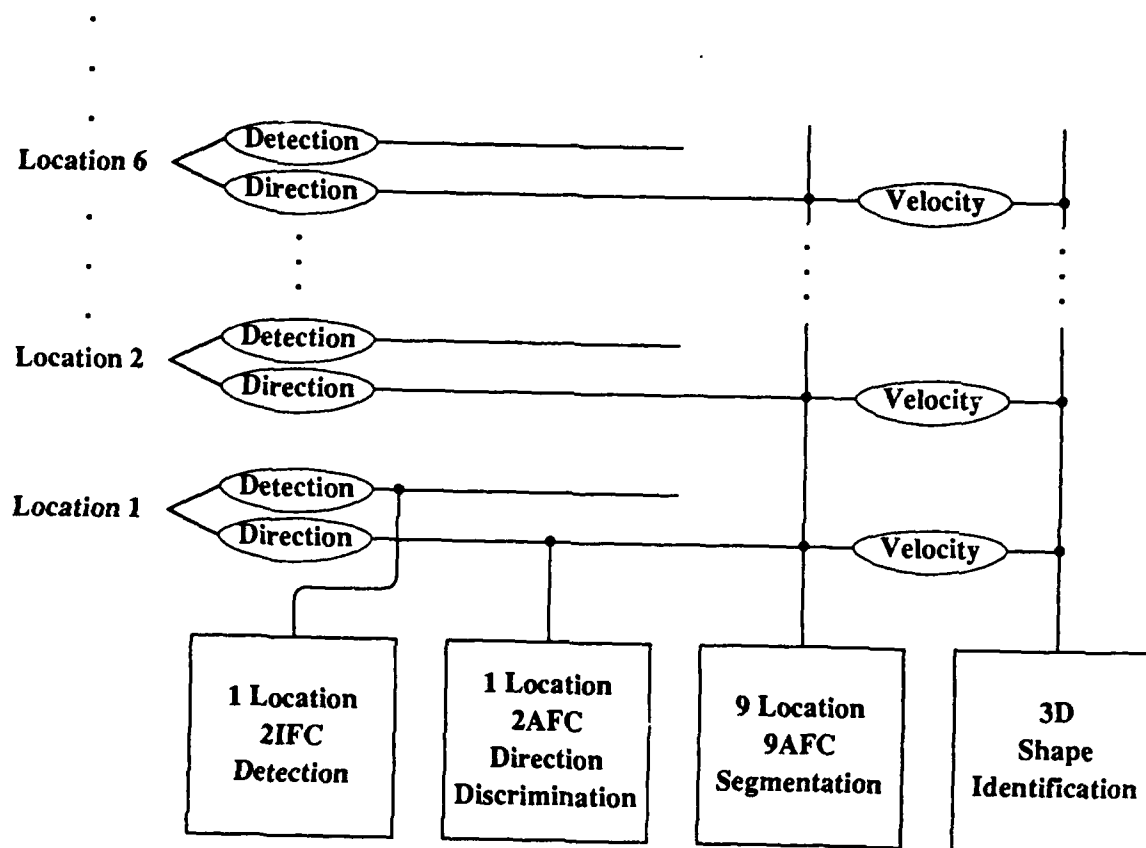


Figure 11. A schematic illustration of the kinds of information required in order to perform each of the experimental tasks. The simple 2IFC detection task may reflect the output of non-motion systems in a single location. The 2AFC discrimination of motion direction task requires the output of a motion direction mechanism in a single location. The 9LFC motion segmentation task requires the output of motion direction mechanisms in a number of locations nearly simultaneously. The 3D shape task requires direction and speed information from a number of locations nearly simultaneously.

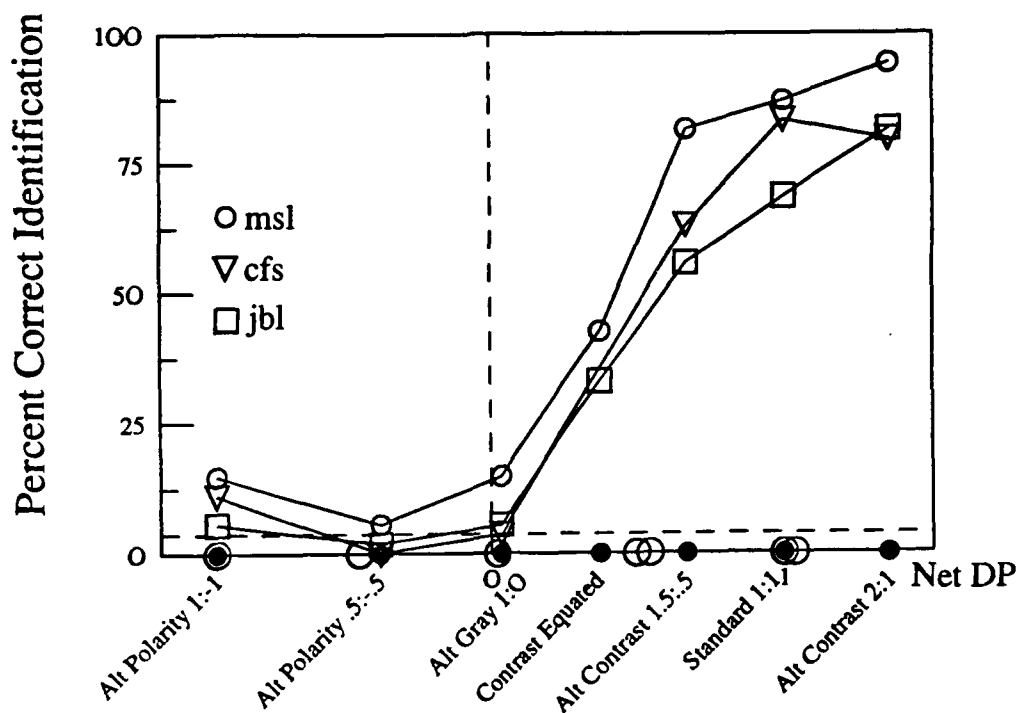


Figure 12. The relation between 3D shape identification performance and computed *net directional power DP* within the window of visibility and above a threshold ϵ . Solid circles on the abscissa are values of *DP* computed from the spectra in Figure 10, panels (b), (d), etc., for an ϵ of $.12 \times$ the maximum power value in the spectrum of the standard stimulus. Open circles on the abscissa are the values of *DP* computed for an ϵ of 0. (The rank order of conditions under the two computations is the same.) The 3D shape identification performance is monotone with *DP* for all reasonable values of $\epsilon \geq 0$.